DOCUMENT RESUME

ED 047 594 48 FL 002 137

AUTHOR	Lado, Robert; And Others
TITLE	Massive Vocabulary Expansion, Phase II: The Effect of Oral and Orthographic Stimuli on the Memorization and Pronunciation of Basic Dialogs. Final Report.
INSTITUTICN	Institute of International Studies (DHEW/OE), Washington, D.C.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-9-7712
PUB DATE	Dec 70
CONTRACT	CEC-0-9-097712-3491
NOTE	118p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$6.58
DESCRIPTORS	*Audiolingual Methods, Basic Skills, Dialogue, Educational Experiments, Evaluation Methods, *Language Instruction, *Language Research, Modern Languages, Second Language Learning, *Spanish, Speech Skills, Statistical Analysis, Student Evaluation, Teaching Methods, Teaching Procedures, *Teaching Techniques, Writing Skills

ABSTRACT

The problems investigated in the present study are the identification of the most effective type of presentation of dialogues in language instruction--auditory, orthographic, or some combination of the two--and the measurement of the accuracy of pronunciation and the amount of dialogue retained by the student. The controlled experiment, performed with a class of first-year college students of Spanish, also examines theory and procedures for teaching writing and reading in the early stages of audiolingual language courses. Major sections contain introductory remarks; experimental procedures; results; summary, conclusions, and recommendations; and a bibliography. Appendixes include various experimental materials. A statistical analysis of the language learning experiment concludes the study. (RL)

ED047594

PA-48
BR-9-7712

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FINAL REPORT

Contract No. OEC-0-9-097712-3491

Massive Vocabulary Expansion, Phase II: The Effect of Oral and Orthographic Stimuli on the Memorization and Pronunciation of Basic Dialogs

Robert Lado, Project Director
Frances Aid, Research Assistant
Maria Kruvant, Research Assistant

December 1970

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Institute of International Studies

FL 002 137

ACKNOWLEDGEMENT

Many colleagues and students contributed in various ways to this research. Although it is not possible to mention everyone, we wish to record our debt to the following and our gratitude to all those not mentioned.

Administrative:

James E. Alatis
Jose M. Hernandez
Elizabeth Johansen

Technical Assistance:

William Kruvant
Leslie Hanzely
American Language Institute

Secretarial:

Carmenza Villalobos
Jennifer Butler
Christiane Kodsi
Carol LeClair

The research reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education, under provisions of Public Law 85-864, Title VI, Section 602, as amended. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position of policy.

STAFF

Robert Lado, Project Director
Frances M. Aid, Research Assistant
Maria Kruvant, Research Assistant

CONSULTANTS

John B. Carroll, psychology, improvement of design
Paul Pimsleur, suggestions for improvement
Othmar Winkler, statistics

School of Languages and Linguistics
Georgetown University
Washington, D.C.

TABLE OF CONTENTS

CHAPTER I. INTRODUCTION

1.1	The problem	1
1.2	Objectives	5
1.3	Related research	6

CHAPTER 2. PROCEDURES

2.1	Training sessions	8
2.1.1	Teaching	8
2.1.2	Testing	8
2.2	Dialogs	8
2.3	Pretests and questionnaires	10
2.4	Subjects	11
2.4.1	Criteria of selection	11
2.4.2	Personal histories	11
2.4.3	Aptitude tests	11
2.4.4	Debriefing	12
2.5	Scoring	13
2.5.1	Memory	13
2.5.1.1	Narrow scoring	13
2.5.1.2	Broad scoring	13
2.5.2	Pronunciation	14
2.6	Equipment	17
2.7	Pilot teaching	17

CHAPTER 3. RESULTS 18

CHAPTER 4. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS 23

5. BIBLIOGRAPHY 26

6. APPENDIXES

6.1	Pretests and questionnaires	30
6.1.1	Personal information	30
6.1.2	Memory aptitude	31
6.1.3	Pimsleur Language Aptitude Battery	31
6.1.4	Attitude questionnaire	32
6.1.5	Debriefing questionnaire	35
6.2	Structure of the lessons	36
6.3	Instructions to subjects	37
6.3.1	Type A	37
6.3.2	Type B	37
6.3.3	Type C	37
6.3.4	Type D	38
6.3.5	Type E	38
6.4	Dialogs	39
6.4.1	Dialog 1	40
6.4.1.1	Full Text	40
6.4.1.2	Phrasing, prompting	53
6.4.1.3	Pronunciation score sheet	54
6.4.2	Dialog 2	55
6.4.3	Dialog 3	59
6.4.4	Dialog 4	63
6.4.5	Dialog 5	67

7. STATISTICAL ANALYSIS OF THE LANGUAGE LEARNING EXPERIMENT
by Othmar W. Winkler 75

THE EFFECT OF ORAL AND ORTHOGRAPHIC STIMULI ON THE MEMORIZATION AND PRONUNCIATION OF BASIC DIALOGS

CHAPTER I. INTRODUCTION

1. The Problem

This research was designed to test experimentally whether differences in the type of presentation of basic dialogs result in differences in amount of material memorized and in accuracy of pronunciation. The differences in type of presentation involved the use of listening and reading as separate factors and in combination with each other. The research is not intended either to support or reject the memorization of dialogs as a teaching strategy; it is intended to compare different ways of memorizing them in order to test the effect of oral and orthographic stimuli in second language learning.

One of the conclusions of the Massive Vocabulary Expansion (MVE) research (Lado, Baldwin and Lobo, 1967) indicated that general labels such as audio-lingual and audio-visual are grossly unspecific and ambiguous as descriptive terms for modes of presentation of language learning materials. That research showed significant differences in the effect of such variables as auditory and orthographic presentation separately and in various combinations. Simultaneous presentation of auditory and graphic forms, for example, was superior to auditory or orthographic presentation alone and to nonsimultaneous combinations when the goal was expansion of recognition vocabulary beyond the basic course. It was suggested that similar research on other elements of second language learning might yield significant results. A promising area for this second phase of investigation was the memorization of basic dialogs.

Dialogs for memorization, introduced by the ASTP intensive language courses and materials of World War II, have become an integral part of the majority of current texts for teaching modern languages. It may not be entirely obvious that major differences in presentation of dialogs are hidden under the ambiguous labels of the methods employed. Dialogs are taught in various ways under the mimicry-memorization label or they appear without instructions to teachers or students under the larger umbrella of "the" audio-lingual method. Therefore this research is focused on the question of whether different results can be obtained in teaching dialogs on the basis of overlooked differences in presentation.

A specific issue in this question of defining and specifying procedures in an audio-lingual approach is the use of writing in the early

stages of a basic course. Some advocates of the audio-lingual approach have insisted that all writing should be withheld from the students for as long as six weeks or even a semester. The argument has been that students must learn to hear and speak the language, and that they will hear and speak it better if the written representation is withheld. Brooks (1960) in considering the materials for level I, "irrespective of the age of the student," states,

> "A sustained experience (of weeks or even months) in listening and speaking must precede training in reading and writing. When the last two skills are begun, the learner must at first read and write what he already knows how to say. The ear must for a time guide the eye, as it inevitably does in becoming literate in the mother tongue;" (p. 123)

Rivers (1968) in characterizing the audio-lingual method states,

> "Some audio-lingual experimenters have suggested periods of as long as twenty-four weeks of purely oral work before students see anything in graphic form, although periods of from six to ten weeks are more common. It is believed that this time lag obviates the interference of native-language habits of pronunciation associated with the printed symbols where these are the same for the two languages, and also that it forces the student to concentrate his attention on accurate and thorough learning of the foreign-language material." (p. 47)

and

> "Students who have immediate access to a graphic representation of what they are learning will often depend on it too much and not give sufficient time to practicing until the work is memorized and internalized. For these reasons the best approach appears to be to present all foreign-language material at first in oral form, especially in the elementary sections of the course; to train students in working with this material orally until they can handle it with ease; then to train them with the script, which they may use as a help to clarification and memorization. (p. 48)

and

> "After the student has heard the correct spoken form of the language, he can attempt to reproduce it himself. Until he has learned the correct pronunciation, he cannot

read it as it would be read by a native speaker." (pp. 51-2)

and finally,

> "The optimum time interval separating the stages of aural presentation and oral repetition from reading and writing is still a matter for experimentation, and experienced teachers differ in their considered judgment at this point." (p. 52)

Scherer and Wertheimer (1964) report on the comparison of college students who studied German by the audio-lingual method and a control group that used the grammar translation approach. The A-L students did not see any written material until the thirteenth week of the course.

O'Connor (1960) recommends a prereading period during which the students are to concentrate on audio-lingual practice to the exclusion of reading. The length of this audio-lingual prereading period may vary:

> "In general when the structure of the foreign language differs radically from that of English, so that new language habits frequently conflict with the student's native English habits, the transition to reading and writing should be relatively later in order to concentrate all available time and effort on the indispensable saturation with the sounds and structural patterns of the foreign language through maximum oral practice." (p. 9)

This belief in the necessity to begin with auditory stimuli and to withhold orthographic stimuli is not new. Otto Jespersen (1904) defended it:

> "Perhaps it is worthwhile here to consider the four ways in which it is possible to communicate the material of a foreign language to pupils. Either (1) the teacher may not let them use any writing at all, but give them everything orally; or (2) he may give them the orthography alone; or (3) he may give them orthography and phonetical transcription together; or finally (4) he may give them phonetical transcription alone.
>
> (1) The first way obviously has the advantage that there is no sound-symbol whatever to confuse the clear apprehension of the pupils; it resembles the manner in which a child learns its mother tongue." (pp. 168-69)

Palmer (1921) also favored an initial auditory period and the exclusion of orthography. If the student wants to learn the language the

natural way, he has to train and drill his ears to do the work for which they were intended. The ear-training starts with imitating sounds, then words, then sentences. A certain amount of memorizing must be done from the beginning. At a later stage the student may acquire his sentences by reading instead of listening. But this will only be after he has become proficient in reproducing what he hears.

Lado (1964) states:

> "Principle 1. Speech before Writing. Teach listening and speaking first, reading and writing next."

and

> "Now that the class have heard and pronounced the word and know what it means, they should see it and read it aloud. Interference from the writing system will eventually have to be faced. If it is faced while the memory of the sound is fresh, it can be overcome." (p. 126)

Even though no prereading period is advocated, the clear implication is that the initial impression should come through the ear.

Lado (1971) recommends a technique for dialog memorization using orthographic stimuli as follows:

> "It is an audio-visual technique in which you write the dialog clearly on the board and read it aloud to the class. You then read it again, inviting the class to read it aloud with you. Next you erase one word in each line and read it again with the class as if the word were still there. You erase a second word in each line and read it again with the class. You continue to read it through with the class while erasing more and more of the words until finally the class is saying the entire dialog while looking at a blank board." (Introduction)

This belief combined with the idea that introduction of the written forms would have very detrimental effects on pronunciation is, then, widely held and deeply rooted. Yet Lado had observed that on meeting a stranger one often does not hear the name accurately even when it is pronounced in English. In such cases, we may ask that the name be spelled verbally; whereupon we may suddenly perceive and recognize it clearly. If this is so in a language one knows well, would it not be even more so in a foreign language?

To the arguments against the presentation of writing at an early stage there are also plausible counterarguments from a cognitive theory of learning which emphasizes the cognitive grasp of the behavior to be learned. It can be argued that familiar graphic symbols might actually help students direct their attention consciously to the unfamiliar sound distinctions in the target language with the result that they may hear them more accurately and remember them better. Phonetic symbols might presumably produce the same results if the students are sufficiently familiar with the symbols to benefit from them. This seems to have been implied in the label, "Aids to listening," used in the ASTP language materials.

The absence of experimental evidence to support these different views led to the design of this research comparing five different strategies for the memorization of dialogs in Spanish by speakers of English. Results of Phase I of the MVE research were influential in determining the design of the current experiment. The project was thought to be relevant on both theoretical and practical grounds. In addition to the practical interest of the research for current classroom teaching and study practices, an important theoretical problem would be tested, namely the principle of exclusive audio-lingual presentation for beginners.

Specifically, the variables investigated were auditory and graphic stimuli, separately and combined in three ways, giving five types of presentation as follows:

TYPE A: Auditory
TYPE B: Graphic
TYPE C: Auditory followed by graphic
TYPE D: Graphic followed by auditory
TYPE E: Simultaneous auditory and graphic

Other variables such as the number of repetitions of the models, time of exposure to the graphic stimulus and the manner of response were held constant. Effectiveness of the presentations was determined on the basis of (1) the amount of text memorized in a given time, and (2) accuracy of pronunciation in recall of the text.

1.2 Objectives

The immediate practical objective was to determine experimentally the most effective type or types of presentation of basic dialogs for memorization by beginners, on the basis of both the amount of text mem-

orized and the quality of pronunciation achieved.

A broader objective was to test the widely held belief that orthographic stimuli should be withheld from the beginner in foreign language teaching.

Finally, in the light of current generalities and ambiguities about approaches to language teaching, an objective was to provide experimental data that would contribute to the formulation of more specifically focused hypotheses of second language learning that might support methodological conclusions.

1.3 Related Research

A number of experiments by Krawiec (1946), Dunkel (1948), Kessman (1959), Postman and Rosenzweig (1956), Pimsleur and Bonkowsky (1961), Pimsleur, Sundland, Bonkowski and Mosbert (1964), and Asher (1961), have compared the relative effectiveness of visual versus auditory presentation of verbal learning material. Krawiec, Dunkel and Kessman found significantly better results when stimuli were presented visually. Postman and Rosenzweig, and Asher, found that students who learned visually and relearned aurally achieved superior performance in comparison with the performance of students who learned aurally and relearned visually. Pimsleur and Bonkowsky (1961), on the other hand, reported greater transfer from aural to visual learning than the reverse. In a modified follow-up of the research, Pimsleur et al. (1964) failed to confirm this finding and concluded that audio to visual order is approximately equal to visual to audio in facilitating relearning.

The problem investigated in the present study differed from the above in that it sought to identify the most effective type of presentation, auditory, orthographic, or some combination of the two, and measured not only the amount of material retained, but accuracy of pronunciation as well.

There is a fairly large literature on visual, auditory, and visual-auditory presentation for rote learning in the native language which provides a background of information: Webb and Wallon (1956), Mowbray (1953), King and Madill (1968). The problem of different strategies of memorization of dialogs in a foreign language has not been studied.

S. Earle Richards and Appel (1956) studied the effects of written words in beginning Spanish. The Lado et al. research on Massive Vocabulary Expansion (1967) investigated the effect of the same differences in presentation for the teaching of passive vocabulary to nonbeginners.

Lipton (1969) in an experiment in French at the FLES level compared the progress of gifted fourth grade pupils who learned the material audio-lingually and read it twenty-four hours later, with a control group that did not read the material. The students who read the material a day later outperformed the controlled group in auditory comprehension.

The present study differs from the above in that it tests the effect of auditory and orthographic stimuli when used in the same lesson alone or in combination with each other. The subjects were college students studying Spanish for the first time, and the experimental effects measured were both amount of material memorized and accuracy of pronunciation when speaking from memory.

CHAPTER 2. PROCEDURES

2.1 Training Sessions

2.1.1 Teaching

The subjects were instructed to attempt to memorize as much as possible of a dialog during each learning session. They were told that they would be tested at the end of the period, and that their tests would be scored on the amount of material memorized and the accuracy of pronunciation. They were directed to repeat aloud during the part of the lesson designed for active practice. The general plan of the lesson was outlined for them.

Each lesson consisted of an introduction of the dialog with English equivalents, followed by an active practice section in which first phrases, then full lines were presented for repetition (See Appendix 6.3). The subjects' responses were recorded during their practice. The lessons averaged 19 minutes, with slight variations in time, ranging from 18'35" to 19'45", due to differing length of the dialogs.

2.1.2 Testing

At the end of each lesson the subjects were immediately tested: half were tested first in writing, then orally; the other half in the opposite sequence. They were asked to either speak or write whatever they could remember of the dialog. For both oral and written tests the subjects attempted to recall the dialog first unprompted, then prompted. To prompt the subjects, one word was provided orally from each line of the dialogs, chosen subjectively on the basis of content value to help recall (See Appendix 6.4.1.2).

2.2 Dialogs

The five dialogs prepared for the experiment were considered appropriate for a beginning course in Spanish with the exception that they were longer than would ordinarily be recommended for classroom use. The extra length of the dialogs was deliberately chosen to make certain that differences in the amount learned were not obliterated by not having enough material to show these differences. The content of each dialog was varied to avoid as far as possible duplication of vocabulary. Some difference in length of the dialogs was permitted.

The text of the dialogs was typed on 5 x 8 cards as they were to be presented in the training sessions. They were then photographed and made into 2 x 2 slides. The lessons were programmed on audio tape recordings according to the respective types of presentation. The timing of the audio and visual sequences was controlled by the manner in which the tapes were recorded, e.g. in one of the combined-stimulus presentations such as Type C (auditory followed by graphic), the time for the S's reading of each phrase or line was equal to the length of time taken to speak each phrase or line (See Appendix 6.5).

The five dialogs were all learned in the same sequence, with the types of presentation randomized according to the system seen in Figure 1. The different types of presentation were rotated among the dialogs and subjects so as to equalize the cumulative effect that learning each successive dialog would have.

Figure 1. Rotation of types of presentation among dialogs and students. (five students per group)

	DIALOG	1	2	3	4	5
GROUP 1	TYPE	A	B	C	D	E
2		E	A	D	C	B
3		B	D	A	E	C
4		D	C	E	B	A
5		C	E	B	A	D
6		A	C	B	E	D
7		E	D	A	B	C
8		D	B	E	C	A

Complete randomization with regard to the types of presentation that would follow each other could not be achieved without altering the randomization of types in serial position, which was assumed to be more important due to the cumulative learning effect of the successive dialogs by the same subject. The actual sequences of the respective types is shown in Figure 2.

The dialogs were taught over a period of two to five weeks. Twenty-nine subjects were taught in the fall of 1969, the remaining eleven were taught in spring 1970.

Figure 2. Sequences of types of presentation.

TYPE	-A	-B	-C	-D	-E
A-	-	AB*	AC	AD*	AE
B-	BA*	-	BC*	BD	BE*
C-	CA	CB*	-	CD	CE*
D-	DA*	DB	DC*	-	DE
E-	EA	EB*	EC*	ED*	-

*Sequences that occurred more than once.

2.3 Pretests and questionnaires.

Pretests were given to the subjects to determine their aptitude for second language learning. These included (1) a memory test parallel to Part V of the Modern Language Aptitude Test (Carroll 1955) prepared by David Horton. (2) The Pimsleur Language Aptitude Battery was administered. These tests identify persons whose aptitude for auditory learning differs from their aptitude for cognitive learning, a difference that was thought might be related to individual differences in performance under the different types of presentation in the experiment. (3) The attitude questionnaire designed for the MVE experiment to evaluate the motivation and attitudes of the subjects toward learning Spanish was also administered. The 7-slot scale of Osgood, Suci and Tannenbaum (1957) for the semantic differential as the index of range and intensity was the basic format. The questions were directed toward Spanish as an academic course, Spanish as an instrument to employment, and Spanish as an integrative force (cf. Lambert et al. 1963). (4) A general questionnaire was given to the subjects to obtain personal histories of previous language contact and foreign language study, as well as personal characteristics such as age, etc. (5) In addition to the testing done at the end of each learning session, a debriefing questionnaire was given to identify the subjects' preferences in learning either auditorily or visually in relation to the five types of presentation of the experiment (see Appendix 6.1.5).

2.4 Subjects.

2.4.1 Criteria of selection

Forty subjects were selected for the experiment on the following criteria:

(a) they were college students;
(b) they were native speakers of English;
(c) they were not natural bilinguals;
(d) they had not studied Spanish previously (many were just beginning a basic course in Spanish).

2.4.2 Personal histories

Of the subjects selected the following information was obtained:

Sex: 25 males, 15 females
Mean age: 19
Academic rank: 25 freshmen
5 sophomores
3 juniors
4 seniors
3 graduate students
Previous formal study of foreign language: Romance languages, average of 4.2 years including Latin, French, and Italian. Nonromance languages, average of 1.0 years including German, Chinese, Russian, Arabic, Greek, and Indonesian.
Previous contact with foreign languages outside formal study: 7 subjects had experience of foreign residence for more than one month, including Cuba (as an infant), France, Germany, Taiwan, and the Philippines. 5 had the experience of some foreign language spoken at home including Tagalog, Polish, French, Dutch, and Indonesian. (see Appendix 6.1.1)

2.4.3 Aptitude tests

Mean scores: Vocabulary memorization: 17 (of possible 24) (Appendix 6.1.2)
Pimsleur Battery, total: 99 (1-117) (Appendix 6.1.3)

Attitude questionnaire: (+3 to -3 on 7-slot scale)

Interest index	- .4
Instrumental index	-1.8
Integrative index	+5

(Appendix 6.1.4)

2.4.4 Debriefing (See Appendix 6.1.5)

Debriefing questionnaires were completed by 24 subjects. The following attitudes and opinions were reported:

22 subjects believed that they learn better by reading than by listening.
2 subjects preferred to learn by listening.

With reference to the five types of presentation in the experiment, the most effective type was held to be:

Type E by 16 S's
Type C by 4 S's
Type B by 3 S's

In addition, 16 considered Type A the least effective. Six subjects commented that Type B was difficult because they felt the need to associate sound with what they read. Three subjects found the simultaneous presentation of Type E distracting or disorientating.

Concerning the learning strategies employed in the attempts to memorize the dialogs, various comments were made. In the following tally of the comments, the numbers in parentheses indicate the number of subjects who reported the respective personal strategies.

Dependence on lexical meaning	(13)
Repetition	(11)
Dependence on cognates	(3)
Attention to syntactic linking	(3)
Dependence on graphic image	(2)
Learning by phrases or units	(2)
Awareness of inflections	(1)
Use of exclamatory forms as cues	(1)

Further comments of the subjects included the following: the futility of hearing a whole dialog twice before the meaning was provided (1); preference for having the graphic presentation on paper rather than on a projected slide

(1); excessive length of the dialogs (2); restrictive shortness of learning session (5); preference for personal interaction with a teacher during learning (2); preference for more repetition of the dialogs (1); admission of low motivation to learn in the experimental situation (1).

2.5 Scoring.

2.5.1 Memory

To score the amount memorized, the data available were the Subjects' written and oral free recall performance, done in every case immediately at the end of a learning session. In scoring, one point was given for each word recalled if it occurred in the original context, i.e. only a minimum of two contiguous words was scored; no points were given for isolated or interrupted words. The subject's score was the total number of contiguous words recalled in each given test situation.

In the interest of finding optimal testing conditions and optimal scoring criteria, several different analyses of the data were made with respect to the amount memorized. The data available provided scores of both oral and written recall, and in each of these, an unprompted and a prompted performance. In addition, for each of these categories, the tests were scored on narrow and broad criteria.

2.5.1.1 For the narrow scoring the following criteria were used:

Written tests: No grammatical errors were accepted. No changes of word order were accepted. Only those spelling errors that clearly reflected interference of English orthographic conventions were accepted, e.g. gemination of consonants.

Oral tests: No grammatical errors were accepted. No changes of word order were accepted. One phoneme distortion per word was accepted unless the word consisted of only one or two segmental phonemes.

In the case of both oral and written tests, "no grammatical error ..." meant that a "one-phoneme distortion" was not accepted if it carried a grammatical function, e.g. an inflectional suffix.

2.5.1.2 For the broad scoring the following criteria were used:

Written tests: Spelling errors indicative of two phonemic distortions were accepted unless the word in question was a one or two segment sequence. In a few cases a larger range of error in long polysyllabic words was accepted by the scorers when it seemed clear that S had remembered the word.

Oral tests: Pronunciation errors up to two phonemic distortions per word were accepted unless it was a word of one or two phonemes. Further distortion was occasionally accepted in the case of long polysyllabic words still recognizable to a native speaker.

In both oral and written data, the following errors were accepted in the broad scoring:

(1) lack of agreement between noun head and modifiers;
(2) lack of agreement between subject and verb;
(3) incorrect verb inflection: tense, mood;
(4) interchange of familiar and formal forms of personal pronouns;
(5) omission of enclitic pronouns from verbal forms;
(6) changes of word order that were syntactically acceptable.

2.5.2 Pronunciation

To score accuracy of pronunciation, the following procedure was followed:

In each dialog approximately thirty-five items were selected, on either phonemic or subphonemic criteria, on the basis of contrastive analysis of English and Spanish. Under consideration were features of stress, vowel and consonant production, consonant clusters, assimilations and possible errors resulting from the interference of the writing system.

Once the items to be scored were selected, certain norms were agreed upon between the judges regarding the range of acceptability in the subjects' pronunciation, i.e. what features of a given item were relevant in considering it right or wrong.

Two judges, a native speaker of Spanish and an American linguist trained in Spanish, scored the oral recall tests for accuracy of pronunciation. In some cases the amount of material recalled was extremely limited, thus data available to judge accuracy of pronunciation was sparse.

The scores were computed as the percentage of correct items figured on the basis of the number of items spoken in a given performance.

There was concern that problems of interpreting the pronunciation scores might arise in relation to the amount of text recalled by the S's.

If they attempted only the easier items, those who recalled more of a dialog might achieve a higher pronunciation score than those who attempted very little. A more realistic measure of their learning might have to take into account the items avoided.

However, an item analysis of the discrete points scored showed no correlation between the number of items attempted and the number correct.

A possible explanation for this might be that cognate words were frequently recalled, but native language interference occasioned frequent pronunciation errors in such words. This hypothesis was not thoroughly investigated.

As a result of the item analysis, it was assumed that the method used to score pronunciation was a valid one.

Figure 3 shows the item analysis done for Dialog 3. The graph is typical of the distribution found for all the dialogs.

Item analysis of pronunciation scores, dialog 3

x = frequency of attempts
y = percent correct

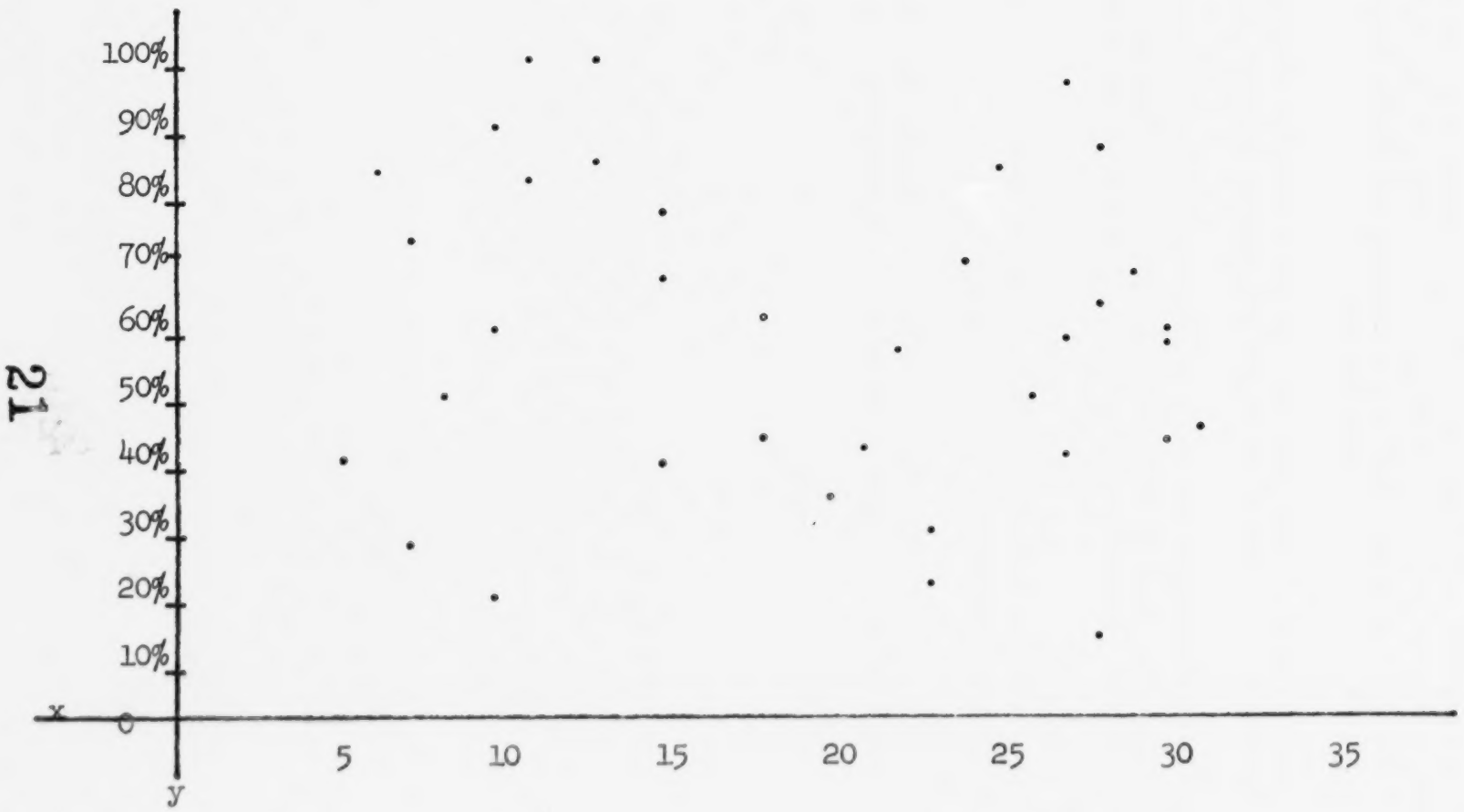


FIGURE 3

2.6 Equipment.

A single lens reflex 35mm. Honeywell Pentax camera with a close-up lens (49mm. - no. 1) and a copy stand with two lights and an art board were used for photographing the texts of the dialogs as they were to be presented in the experiment. The film used was Kodak High Contrast Copy Film M135-36, which was developed commercially. The negatives were then mounted into 2 x 2 cardboard frames by the research team.

Tapes were prepared by the research team in the recording studios of the Georgetown University School of Languages and Linguistics.

The specially designed equipment developed for the MVE research to synchronize presentation of audio and visual stimuli, could not be used as planned in the experiment, because it could not be made to function by the supplier. Use of less sophisticated equipment, therefore, resulted in somewhat less precision in controlling timing and sequencing in the presentation of the dialogs.

The training sessions and the testing was done in the American Language Institute language laboratory at Georgetown. A Kodak Carousel 700 slide projector was used for the visual presentations.

2.7 Pilot Teaching.

Pilot training sessions were held previous to final preparation of material for the experiment. Of concern in the pilot sessions were (1) the length of the dialogs; (2) the timing of audio and visual stimuli; (3) the number of repetitions of phrases by the S's during learning; (4) the selection of words for prompting during the testing.

Results showed that (1) the dialogs were of sufficient length and difficulty to serve the purposes of the experiment. They were, of course, intentionally made longer than what is recommended for classroom use so that in the experimental situation the optimal performance of any subject could be observed. The range of scores among the S's proved the chosen length of the dialogs to be realistic for this experiment. (2) The timing of the audio and visual stimuli was decided upon as described in Appendix 5. (3) Two repetitions of each phrase by the S's were thought to be sufficient. (4) A word from each line of the dialogs that was generally most effective was chosen for prompting on the basis of lexical meaning, i.e. the items most suggestive of the content of each line.

CHAPTER 3. RESULTS

3.0 A detailed statistical discussion of the data yielded by the experiment is found in the appended report by the consulting statistician. The findings are reported here.

3.1 The first objective of the experiment was to measure the differences in the amount of text recalled by S's after learning under the five types of presentation.

Among the various sets of measures of recall obtained for each Subject, those used for detailed statistical analysis were the written, narrowly scored, unprompted tests (WNU). This set of scores was chosen after preliminary investigations showed high correlation among the different sets of scores (Table 18) and gave the same results with respect to the rank of the five types of presentation. That is, although prompting in many cases produced higher memory scores, they were proportionally the same with respect to each type of presentation. The oral test, which yielded slightly higher scores on memorization, are reported in Table A and Figure 4. They parallel the WNU scores quite closely. In addition, although a broad scoring in which a greater range of error was permitted yielded higher scores, the scoring procedure did not alter the findings relative to the types. Since the most complete data were available among the written scores, these were used for further analysis concerned with amount of text memorized.

Table A and Figure 4 show a comparison of the five types of presentation in terms of the arithmetic mean of the scores achieved under each type. The greatest amount of recall of the dialogs by both WNU and ONU was shown by Type B, graphic only--37.2 and 39.22 items respectively. The second largest amount of recall occurred with Type E, simultaneous graphic and auditory presentation--WNU 33.08 and ONU 33.77. The least amount recalled occurred after learning by Type A, auditory only--WNU 19.57 and ONU 22.37.

The differences in amounts recalled are large: WNU 37.2, Type B, is 90% larger than the weakest, WNU 19.57, Type A. The second largest, WNU 33.08, Type E, is 67% larger than Type A. The differences are of approximately the same magnitude whether we use the WNU or the ONU arithmetic means. The analysis of variance showed the differences among the mean scores to be highly significant, with an F-ratio of 15.13.

The results in all the types which used graphic presentation alone or in combination with oral presentation were higher than oral presentation alone. The differences are significant statistically.

It is clear that these college students with considerable experience in the study of other foreign languages memorized the dialog material more efficiently when they could see it in written form either alone or in connection with hearing it.

Table A. Arithmetic mean of amount of text memorized by five types of presentation.

		WNU	ONU
Type	A	19.56	22.37
	B	37.2	39.22
	C	29	30.06
	D	29.50	31.82
	E	33.08	33.77

Mean scores for types of presentation

ONU_____

WNU----------

All dates included

A - auditory only
B - graphic only
C - auditory + graphic
D - graphic + auditory
E - simultaneous

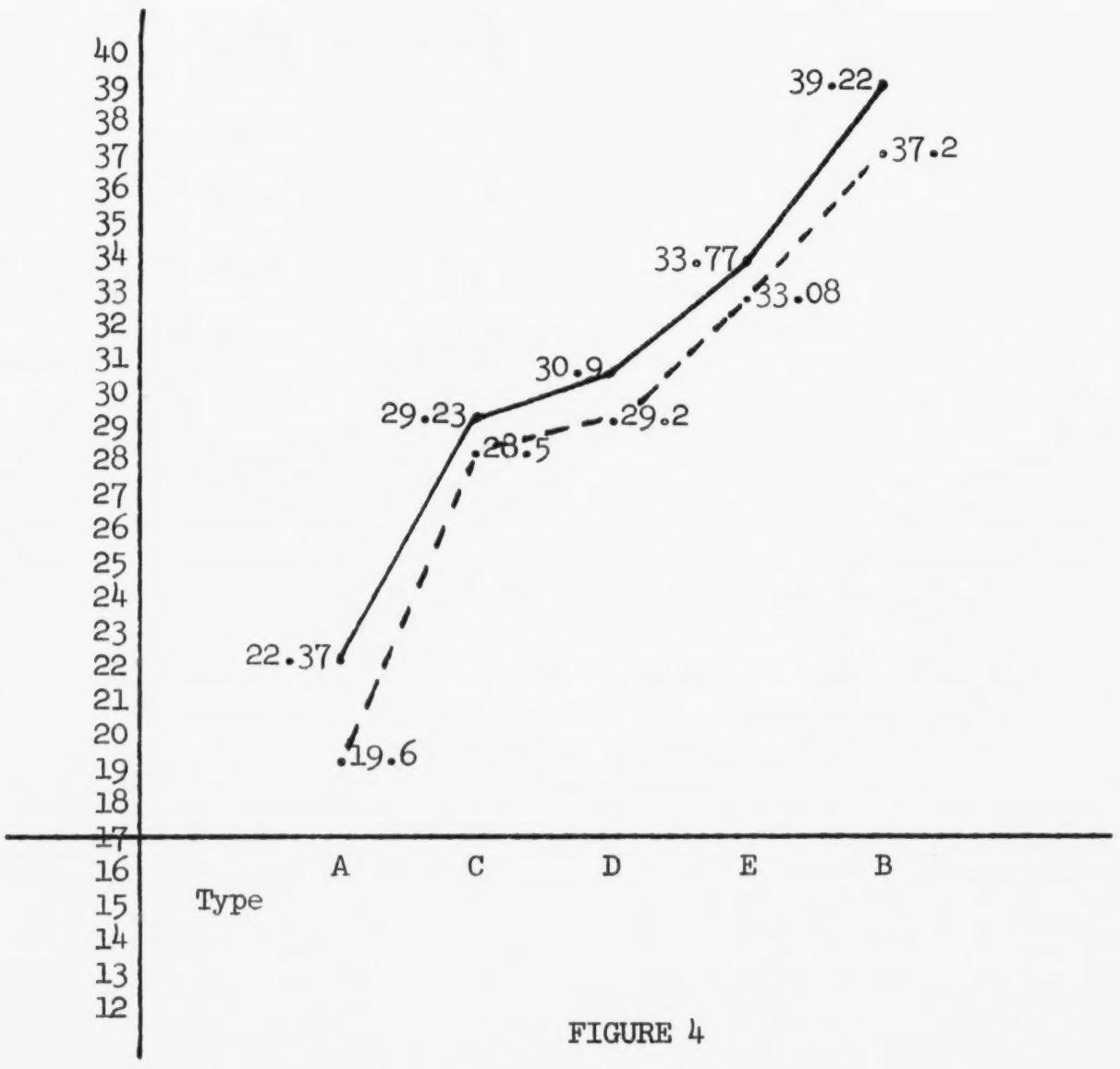


FIGURE 4

3.2 The second focus of the experiment was accuracy of pronunciation in the S's performance after learning by the respective types of presentation. This was crucial because of the argument that exposure to writing in the early stages of language teaching will have a strong negative effect on pronunciation.

Table B and Figure 5 show the percentages of total correct items over the total attempted among the five types of presentation.

Type A, auditory only, shows 62% correct items in pronunciation and Type E, 60%. The difference is not statistically significant. In fact, the differences in pronunciation among all five types are not significant. The actual percentages are 57% for Type C, oral followed by graphic; 56% for Type B, graphic only; and 55% for Type D, graphic followed by oral.

Summarizing the results, we see that exposure to the written form of the dialogs had a statistically significant favorable effect on the amount of material memorized and it did not have a statistically significant negative effect on pronunciation for this experiment. It is important to point out that the students had had considerable experience in the study of foreign languages, an average of 3.7 years per S.

Table B. Mean pronunciation scores. Number of subjects, 37.

Type	Percent correct
A	62%
B	56%
C	57%
D	55%
E	60%

Differences not statistically significant.

Pronunciation scores: percentage correct

37 students - percentage of total correct/total attempts

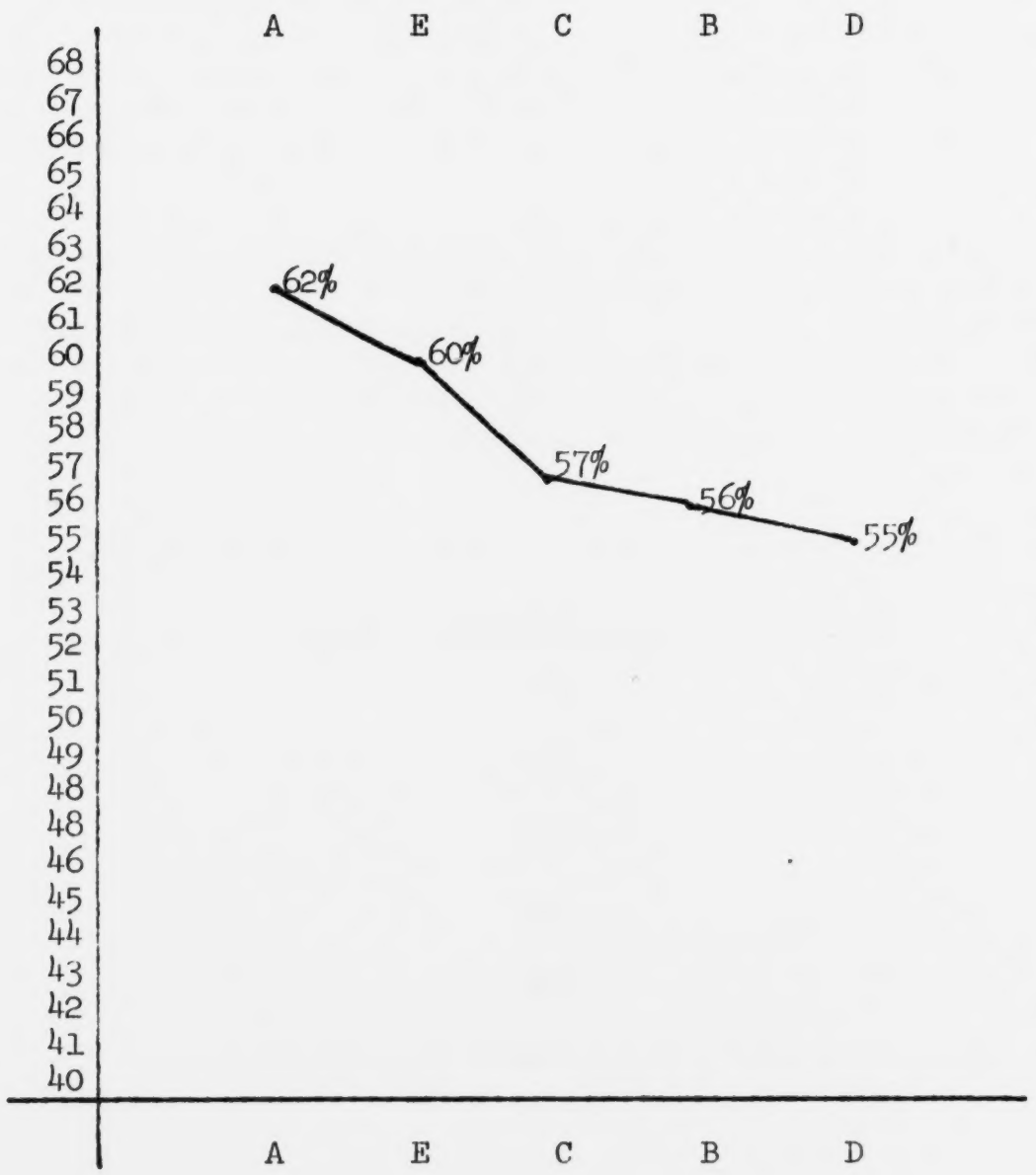


FIGURE 5

CHAPTER 4. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Five types of presentation were used in rotation among five different dialogs that were presented to 37 S's for memorization. Their oral and written recall tests after each learning session were analyzed in terms of the amount of text recalled and accuracy of pronunciation. The recall tests showed the five types of presentation to be effective in the following descending order:

(1) Type B Graphic
(2) Type E Simultaneous auditory-graphic
(3) Type D Graphic + auditory
(4) Type C Auditory + graphic
(5) Type A Auditory

Differences among the five types were found to be highly significant.

Measures of pronunciation showed the types of presentation to rank as follows:

(1) Type A Auditory
(2) Type E Simultaneous auditory-graphic
(3) Type B Graphic
(4) Type C Auditory followed by graphic
(5) Type D Graphic followed by auditory

We are well aware of the limitations of our findings and the caution with which we must report them. Is the strong positive effect of the orthographic presentation typical of college students in general or is it restricted to these with considerable experience in studying other foreign languages? Is the regularity of Spanish spelling responsible for the absence of significant negative effect on pronunciation, and if so would there be more negative effects in the case of French or German?

Of course, Spanish spelling is regular for a literate Spanish speaker. A literate English speaker who reads Spanish spelling for the first time has no way of knowing what the regular pronunciation is going to be. Hence he is likely to be misled by the spelling. On the other hand, since S's had studied other foreign languages, including French in many instances and Italian and Latin in others, they would presumably be able to make intelligent guesses at the sounds of the words on the basis of their written forms.

In spite of these considerations, we were not prepared for the magnitude of the differences.

For the past four years the senior investigator has been directing

his students to explore the relative effect of graphic and auditory stimuli alone and combined in teaching French, German, Russian, Portuguese and English as a foreign or second language at the college, high school and FLES levels. They are usually pilot studies without full rigorous control, but their cumulative effect seems to be in the direction of the positive effect of graphic stimuli combined with auditory models, at least with well qualified students. These dimensions will have to be tested under rigorously controlled conditions, but it does not seem entirely unjustified to state that our findings have challenged the need for a prereading all oral beginning period at least at the college level. For Spanish, they seriously challenge the necessity for presenting material first in an exclusively oral modality.

If the results are replicated under differing conditions, we might want to modify the classic ordering, hearing, speaking, reading, writing in the learning of foreign language skills to hearing-seeing, speaking, reading, writing.

The research on Massive Vocabulary Expansion showed that simultaneous hearing and reading of new words was more effective than hearing only or reading only in passive recognition of vocabulary beyond the basic course. The present study seems to show that the graphic stimuli are also helpful at the beginning level.

It is also noted that the effect of reading and of hearing are not cumulative, that is, one is not simply added to the effect of the other. Language learning is not an amorphous matter such that by merely providing a multisensory input we obtain cumulatively better results.

Finally, it should be of more than passing interest that the opinions of the S's in the debriefing questionnaire show considerable, though not complete, agreement between preferences and test results:

Type E Simultaneous was preferred by most subjects--16 out of 24 questioned. They believed that they learned better by this type of presentation. Only three subjects preferred Type B, even though this method produced the largest average amount of recall. Six subjects commented that Type B was difficult because they felt the need to associate sound with what they heard. Three subjects considered Type E distracting. Sixteen subjects considered Type A the least effective, and so it was.

Recommendations

We urge that the research be extended to other languages and to high school students as well. The senior investigator is directing a

parallel Master's degree thesis using German as the foreign language. The design can be simplified to include only the three types, auditory, written, and auditory-written simultaneously. It might well be simplified further to auditory versus auditory-written only.

5. BIBLIOGRAPHY

Asher, James J. 1961. Sensory interrelationships in the automated teaching of foreign languages. Research supported by Grant No. 7-04-091, United States Office of Education, Department of Health, Education and Welfare.

Brent, Sandor B. 1969. Linguistic unity list length, and rate of presentation in serial anticipation learning. Journal of verbal learning and verbal behavior. 8. 7079.

Brooks, Nelson. 1964. Language and language learning--theory and practice. 2d ed. New York, Harcourt, Brace and Company. 1st ed. 1960.

Bruner, J. S. 1964. The course of cognitive growth. American psychologist, 19. 1-15.

Carroll, John B. 1961. Research in teaching foreign languages. Ann Arbor. Also in N. Gage, ed., 1963. Handbook of research on teaching. Chicago, Rand McNally. 1060-1100.

_____ and Stanley M. Sapon. 1955. Modern languages aptitude test. New York, Psychological corporation.

_____. 1966. Research in foreign language teaching; the last five years. In Mead, Robert G., ed. Reports of the working committees of the northeast conference. 7-58. New York, MLA material center.

Chastain, Kenneth D. 1968a. A comparison of the audio-lingual habit theory and the cognitive code-learning theory to the teaching of introductory college Spanish. Dissertation abstracts, 29. 830A.

_____ and Frank J. Woerdehoff. 1968b. A methodological study comparing the audio-lingual habit theory and cognitive code-learning theory. Modern language journal. 2. 268-279.

Dunkel, H. B. 1948. Second language learning. Boston, Ginn.

Fries, Charles C. 1945. Teaching and learning English as a foreign language. Ann Arbor, University of Michigan Press.

Hinz, Marian. 1969. Effect of response mode on learning efficiency. Audio-visual communication review. 17. 77-83.

Jespersen, Otto. 1904. How to teach a foreign language. London, George Allen and Unwin Ltd.

Johnson, Ronald E. 1970. Recall of prose as a function of the structural importance of the linguistic units. Journal of verbal learning and verbal behavior. 9. 12-20.

Kessman, M. 1959. Experimental comparison of writing-speaking versus speaking-writing sequences in learning paired-associates. Unpublished doctoral dissertation, Indiana University.

King, David J., and Alan W. Lau. 1963. A comparison of three scaling techniques in estimating the accuracy of written recall. Journal of general psychology. 69. 203-207.

_____ and R. Bruce Harper. 1967. Scaling the accuracy of recalls of stories under restricted conditions of judgment. Psychological record. 17. 531-535.

_____ and Jon Madill. 1968. Complex methods of presentation, internal consistency of learning material, and accuracy of written recall. Psychological reports. 22. 777-782.

Kjeldegaard, P. M. 1962a. Commonality scores under instructions to give opposites. Psychological report. 11. 219-220.

_____. 1962b. Variables affection recall in one-trial learning. Paper read at Eastern Psychological Association, Atlantic City.

_____. 1962c. Predicting paired-associate learning speed. Psychological report. 11. 353-354.

Krawiec, R. S. 1946. A comparison of learning and retention of materials presented visually and auditorally. Journal of general psychology. 34. 179-195.

Lado, Robert. 1964a. Language testing. New York, McGraw-Hill.

_____. 1964b. Language teaching, a scientific approach. New York, McGraw-Hill.

_____. 1971. Teachers' manual for Lado English series. New York, Simon and Schuster Inc.

_____, B. Baldwin, and F. Lobo. 1967. Massive vocabulary expansion in a foreign language beyond the basic course: the effects of stimuli timing and order of presentation. (Final Report Project No. 5-1095, Grant No. OE-6-14-021), United States Department of Health, Education and Welfare, Office of Education, Bureau of Research.

_____, B. Baldwin, and F. Lobo. Article in the Florida language newsletter or reporter.

_____ and Charles C. Fries. 1954. English pronunciation. Ann Arbor, University of Michigan Press.

Lambert, Wallace E., R. C. Gardner, H. C. Barid, and K. Tunstall. 1963. Attitudinal and cognitive aspects of intensive study of a second language. Journal of abnormal and social psychology. 4. 358-368.

Levitt, Eugene E. 1956. A methodological study of the presentation of connected verbal stimuli for quantitative memory experiments. Journal of experimental psychology. 52. 33-38.

Martin, Edwin and Donald A. Walters. 1969. Subject uncertainty and word-class effects in short-term memory for sentences. Journal of experimental psychology. 80. 47-51.

Mowbray, G. H. 1953. Simultaneous vision and audition: the comprehension of prose passages with varying levels of difficulty. Journal of experimental psychology. 46. 365-372.

Mueller, Theodore H. and Henry Niedzielski. 1968. The influence of discrimination training on pronunciation. Modern language journal. 52. 410-416.

Muller, Daniel. 1965. The effect upon pronunciation and intonation of early exposure to the written word. Modern language journal. 49. 411-413.

_____ and Muller. 1968. The problem of interference in beginning Portuguese. Modern language journal. 52. 201-505.

O'Connor, Patricia. 1960. Modern foreign languages in high school: prereading instruction. Washington, United States Department of Health, Education, and Welfare, Office of Education.

Osgood, D. E., C. J. Suci, and P. H. Tannebaum. 1957. The measurement of meaning. Urbana, University of Illinois Press.

Palmer, Harold E. 1921. The principles of language study.

Pimsleur, Paul and R. J. Bonkowski. 1961. Transfer of verbal material across sense modalities. Journal of educational psychology. 52. 104-107.

_____, D. M. Sundland, R. J. Bonkowski, and L. Mosberg. 1964. Further study of the transfer of verbal materials across sense modalities. Journal of educational psychology. 55. 96-102.

_____. 1966. Pimsleur language aptitude battery. New York, Harcourt, Brace and World.

Postman, Leo and M. Rosenzweig. 1956. Practice and transfer in the visual and auditory recognition of verbal stimuli. American journal of psychology. 69. 209-226.

Richards, S. Earle. 1956. The effects of written words in beginning Spanish. Modern language journal. 40. 129-226.

Rivers, Wilga M. 1968. Teaching foreign-language skills. Chicago, University of Chicago Press.

Rojas, Pauline M. 1952. Preface to Charles C. Fries, American English series. Boston, D. C. Heath.

Scherer, George A. C. and Michael Wertheimer. 1964. A psycholinguistic experiment in foreign-language teaching. New York, McGraw-Hill.

Sherman, Martin F. and M. T. Turvey. 1969. Modality differences in short term serial memory as a function of presentation rate. Journal of experimental psychology. 80. 335-338.

Wanat, Stanley and Harry Levin. 1967. Studies of oral reading: XI. The eye-voice span: reading efficiency and syntactic predictability. United States Office of Education.

Wardhaugh, Ronald. 1970. An evaluative comparison of present methods for teaching English phonology. TESOL quarterly. 4. 63-73.

Webb, W. B. and E. J. Wallon. 1956. Comprehension by reading versus hearing. Journal of applied psychology. 40. 237-240.

6. APPENDIXES

6.1. Pretests and questionnaires.

6.1.1. Personal information.

This information was obtained prior to final selection of subjects for the experiment.

Name _____

Age _____

Sex _____

School _____

Class _____

Semester in college _____

Native language _____

Other language at home _____

Language (s) studied in school _____how long_____

_____ _____

_____ _____

other contact with foreign language _____

6.1.2. Memory aptitude.

A memory test, parallel to Part V of the Modern Language Aptitude Test, was administered to all subjects. It was prepared by David L. Horton of the University of Kentucky and secured through Dr. John B. Carroll.

Horton's adaptation differs from the MLA form in that all the instructions are given at the beginning rather than separately. The student is asked to read them and indicate that he understands them.

Kjeldergaard's shortened times (three minutes to take the test rather than the standard eight minutes) were employed, because they have proved to provide a better correlation between the test and paired-associate tasks (Kjeldergaard 1962).

6.1.3. Pimsleur Language Aptitude Battery (1966).

Form S, was administered to all subjects.

6.1.4. Attitude questionnaire.

Please print:

Name: _____ Date: _____
(Last First Middle)

Instructions: Place an "x" in one of the marked-off slots to indicate your position with regard to the statements below. Work rapidly but carefully. We are interested in honest,sincere responses. Your answers will not be seen by anyone other than the staff of the research project.

Example:	I think English is beautiful.	_	_	x	_	_	I think English is ugly.

Begin:

1.	Spanish is my favorite course.	__	__	__	__	__	Spanish is my least favorite course.
2.	I prefer Spanish to any other foreign language.	__	__	__	__	__	I prefer any other foreign language to Spanish.
3.	I have a facility for Spanish.	__	__	__	__	__	I don't have a facility for Spanish.
4.	I spend a lot of time studying Spanish.	__	__	__	__	__	I don't spend a lot of time studying Spanish.
5.	I try to use Spanish whenever I can.	__	__	__	__	__	I never try to use Spanish unless I have to.
6.	I could not fulfill my professional aspirations without studying Spanish.	__	__	__	__	__	I could fulfill my professional aspirations without studying Spanish.

	most of all						least of all
a. I would like to travel in Spanish-speaking countries.	__	__	__	__	__	__	__
b. I enjoy conversing with Spanish-speaking people.	__	__	__	__	__	__	__
c. I would enjoy reading Spanish newspapers, journals, and literature.	__	__	__	__	__	__	__
d. Spanish will allow me to meet more and varied people.	__	__	__	__	__	__	__
e. I would like to live in a Spanish-speaking country for a year or more.	__	__	__	__	__	__	__
f. Study of a foreign language makes one a more broad-minded, enlightened individual.	__	__	__	__	__	__	__
g. Contact with Spanish-speaking people and their cultures is more important to me than contact with any other foreign language group.	__	__	__	__	__	__	__
h. Contact with Spanish-speaking people and their cultures is more important to me than contact with my native language group.	__	__	__	__	__	__	__

		most of all						least of all
a.	I would like to teach Spanish.	__	__	__	__	__	__	__
b.	I would like to use Spanish as an interpreter or translator.	__	__	__	__	__	__	__
c.	I would like to teach Spanish, literature, civilization, or linguistics.	__	__	__	__	__	__	__
d.	I would like to use it in my profession or job, which is not a., b., or c., above.	__	__	__	__	__	__	__
e.	Spanish is required for my degree.	__	__	__	__	__	__	__
f.	I enjoy Spanish for its own sake.	__	__	__	__	__	__	__
g.	I enjoy learning in general.	__	__	__	__	__	__	__
h.	I would like to continue mu study of Spanish after this course, even if not required.	__	__	__	__	__	__	__

7.	Knowing Spanish will help me better understand the culture and mentality of Spanish-speaking people	__	__	__	__	__	Knowing Spanish will not help me understand their culture and mentality.

6.1.5. Debriefing questionnaire.

Please answer the following questions.

1) Do you consider that you learn better, a) aurally (by listening) or b) visually (by reading)?

_____

2) What strategies did you use in memorizing the dialogues? (Please indicate any variations under the specific type of presentations).

_____

_____

_____

_____

_____

3) What is your opinion concerning the effectiveness of the different types of presentation? listening, reading, listening reading, reading listening, simultaneous reading and listening.

_____

_____

_____

_____

If you have any further insights please write them now or bring your comments to our office.

_____

_____

_____

_____

6.2. Structure of the lessons.

(1) Denotation.
   (a) Whole dialog presented twice;
   English equivalent spoken once.
   (b) Each successive line of the dialog presented twice, followed by its English equivalent.

(2) Practice.
   (a) Each line divided into phrases: if a line x is divided into phrases 1, 2, 3, etc., 1 is presented for repetition, then 2, then 3, etc., then x.
   (b) Each line as a whole is presented again twice; S's repeat.

6.3. Instructions to subjects.

The following instructions were included on the tape recordings at the beginning of each of the S's five lessons in the experiment.

6.3.1. Type A, auditory presentation.

In this learning session, you will attempt to memorize a dialog in Spanish.

The dialog will be presented in the following manner:

You will hear the entire dialog spoken twice in Spanish. It will then be spoken in English.

Then each separate line of the dialog will be spoken in Spanish twice, followed by its equivalent in English.

Next you will hear each line of the dialog broken into phrases for your repetition and practice. Each phrase and line will be said twice followed by a pause for your repetition.

After the dialog has been practiced in this way, it will be reviewed line by line for your repetition. Each line will be spoken twice.

6.3.2. Type B, graphic presentation.

In this learning session, you will attempt to memorize a dialog in Spanish.

The dialog will be presented in the following manner:

You will read the entire dialog in Spanish as it is presented (twice) on the screen. It will then be spoken in English.

Then each separate line will be read on the screen in Spanish. Each line will be left on the screen long enough to be read twice. Then you will hear the equivalent in English.

Next you will read each line of the dialog broken into phrases for your oral practice. Read each phrase or line to yourself twice. After the projected stimulus is turned off, say the phrase aloud.

After the dialog has been practiced in this way, it will be reviewed line by line for your oral practice. Remember that you are to speak <u>after</u> your silent reading of each line.

Now begin to read.

6.3.3. Type C, auditory graphic presentation.

In this learning session, you will attempt to memorize a dialog in Spanish.

The dialog will be presented in the following manner:

You will hear the entire dialog spoken in Spanish. Then you will read it silently in Spanish. It will then be spoken in English.

Next you will hear each line spoken separately in Spanish, then read it on the screen. Its equivalent will be spoken for you in English.

Then you will succesively hear, then read, each line of the dialog broken into phrases for your repetition and practice. Each phrase will be said twice, then projected for your silent reading, followed by a pause for your repetition.

After the dialog has been practiced in this way, it will be reviewed line by line for your repetition. Remember to repeat after the visual stimulus is turned off.

6.3.4. Type D, graphic auditory presentation.

In this learning session, you will attemp to memorize a dialog in Spanish.

The dialog will be presented in the following manner:

You will first read the entire dialog in Spanish. Then the dialog will be spoken in Spanish followed by its equivalent in English.

Then you will see each line of the dialog on the screen; then hear it spoken in Spanish. It will be said in English.

Next you will succesively see, then hear, each line of the dialog broken into phrases for your repetition and practice. Each phrase will be presented visually, then said twice, followed by a pause for your repetition.

After the dialog has been practiced in this way, it will be reviewed line by line for your practice.

Now begin to read.

6.3.5. Type E, simultaneous auditory and graphic presentation.

In this learning session, you will attempt to memorize a dialog in Spanish.

The dialog will be presented in the following manner:

You will hear the entire dialog spoken twice in Spanish. At the same time you will read it silently in Spanish. It will then be spoken in English.

Then each separate line of the dialog will be said in Spanish as you read it on the projection screen, after which you will hear its equivalent in English.

Next you will see and hear each line of the dialog broken into phrases for your repetition and practice. Each phrase will be said twice while you see it, followed by a pause for your repetition.

After the dialog has been practiced in this way, it will be reviewed line by line for your repetition. As you read the lines, each will be spoken twice.

6.4. Dialogs.

The full text of a dialog as presented in a teaching session in the experiment is given for Dialog 1 with Type A. For all other dialogs the text, the English equivalent, the phrase divisions (marked off by slant bars), the words used for prompting (underlined), and the items selected to be scored for pronunciation are given without repeating the full text of a lesson.

6.4.1. Dialog 1: full text with Type A presentation.

6.4.1.1. (Instructions as in 6.3.1.)
S's hear the entire dialog spoken twice.

-- ¿Qué te pasó ayer? No viniste a la Universidad.

-- No pude. Tuve un mal día. Estuve enfermo.

-- ¡Qué lástima! La clase fue estupenda, y además no tuvimos el examen.

-- ¿Por qué no? ¿De qué habló el profesor?

-- Tampoco vino él. Fue un amigo suyo, un artista.

-- ¿Fue el profesor de arte indígena que acaba de llegar de México?

-- Sí, nos dio una conferencia sobre el arte azteca.

-- ¿Cómo se llama este señor?

-- Es el profesor Emilio Gonzalez. Conoces a su sobrina que está
en la misma clase.

-- ¿María del Carmen? Pero ella no es mexicana.

-- No, la familia es de Panamá, pero don Emilio es profesor en
la Universidad de México.

-- Bueno, como no estuve en la clase, voy a pedir a María del Carmen
que me explique el arte azteca.

(The S's always hear the English equivalent spoken once in English.)

-- What happened to you yesterday? You didn't come to school.

-- I couldn't. I had a bad day; I was sick.

-- Too bad! The class was great, and besides, we didn't have the exam.

-- Why not? What did the professor talk about?

-- He didn't come either. It was a friend of his, an artist.

-- Was it the professor of Indian art that just arrived from Mexico?

-- Yes, he gave us a lecture on Aztec art.

-- What is his name?

-- Professor Emilio González. You know his niece who is in that same class.

-- María del Carmen? But she isn't Mexican.

-- No, the family is from Panama, but Don Emilio is a professor at the University of Mexico.

-- Well, since I wasn't at class, I'll ask María del Carmen to explain Aztec art to me.

(The student hears each separate line of the dialog spoken in Spanish twice, followed by its equivalent in English.)

-- ¿Qué te pasó ayer? No viniste a la Universidad.

¿Qué te pasó ayer? No viniste a la Universidad.

What happened to you yesterday? You didn't come to school.

-- No pude. Tuve un mal día. Estuve enfermo.

No pude. Tuve un mal día. Estuve enfermo.

I couldn't. I had a bad day. I was sick.

-- ¡Qué lástima! La clase fue estupenda, y además no tuvimos el examen.

¡Qué lástima! La clase fue estupenda, y además no tuvimos el examen.

Too bad! The class was great, and besides, we didn't have the exam.

-- ¿Por qué no? ¿De qué habló el profesor?

¿Por qué no? ¿De qué habló el profesor?

Why not? What did the professor talk about?

-- Tampoco vino él. Fue un amigo suyo, un artista.

Tampoco vino él. Fue un amigo suyo, un artista.

He didn't come either. It was a friend of his, an artist.

-- ¿Fue el profesor de arte indígena que acaba de llegar de México?

¿Fue el profesor de arte indígena que acaba de llegar de México?

Was it the professor of Indian art that just arrived from Mexico?

-- Sí, nos dio una conferencia sobre el arte azteca.

Sí, nos dio una conferencia sobre el arte azteca.

Yes, he gave us a lecture on Aztec art.

-- ¿Cómo se llama ese señor?

¿Cómo se llama ese señor?

What is his name?

-- Es el profesor Emilio González. Conoces a su sobrina que está en la misma clase.

Es el profesor Emilio González. Conoces a su sobrina que está en la misma clase.

Professor Emilio González. You know his niece who is in the same class.

-- ¿María del Carmen? Pero ella no es mexicana.

¿María del Carmen? Pero ella no es mexicana.

María del Carmen? But she isn't Mexican.

-- No, la familia es de Panamá, pero don Emilio es profesor en la Universidad de México.

No, la familia es de Panamá, pero don Emilio es profesor en la Universidad de México.

No, the family is from Panama, but Don Emilio is a professor at the University of Mexico.

-- Bueno, como mo estuve en la clase, voy a pedir a María del Carmen
que me explique el arte azteca.
Bueno, como no estuve en la clase, voy a pedir a María del Carmen
que me explique el arte azteca.
Well, since I wasn't in class, I'll ask María del Carmen to explain Aztec
art to me.

(The student will hear each phrase and line said twice followed by a pause for repetition.)

-- ¿Qué te pasó ayer?

¿Qué te pasó ayer? (repetition)

No viniste

No viniste (repetition)

a la Universidad.

a la Universidad. (repetition)

¿Qué te pasó ayer? No viniste a la Universidad.

¿Qué te pasó ayer? No viniste a la Universidad. (repetition)

-- No pude.

No pude.

Tuve un mal día.

Tuve un mal día.

Estuve enfermo.

Estuve enfermo.

No pude. Tuve un mal día. Estuve enfermo.

No pude. Tuve un mal día. Estuve enfermo.

-- ¡Qué lástima!

¡Qué lástima!

La clase

La clase

fue estupenda,

fue estupenda,

y además

y además

no tuvimos el examen.

no tuvimos el examen.

¡Qué lástima! La clase fue estupenda, y además no tuvimos el examen.

¡Qué lástima! La clase fue estupenda, y además no tuvimos el examen.

-- ¿Por qué no?

¿Por qué no?

¿De qué habló el profesor?

¿De qué habló el profesor?

¿Por qué no? ¿De qué habló el profesor?

¿Por qué no? ¿De qué habló el profesor?

-- Tampoco vino él.

Tampoco vino él.

Fue un amigo suyo,

Fue un amigo suyo,

un artista.

un artista.

Tampoco vino él. Fue un amigo suyo, un artista.

Tampoco vino él. Fue un amigo suyo, un artista.

-- ¿Fue el profesor

¿Fue el profesor

de arte indígena

de arte indígena

que acaba de llegar

que acaba de llegar

de México?

de México?

¿Fue el profesor de arte indígena que acaba de llegar de México?

¿Fue el profesor de arte indígena que acaba de llegar de México?

-- Sí,

Sí,

nos dio una conferencia

nos dio una conferencia

sobre el arte azteca.

sobre el arte azteca.

Sí, nos dio una conferencia sobre el arte azteca.

Sí, nos dio una conferencia sobre el arte azteca.

-- ¿Cómo se llama

¿Cómo se llama

este señor?

este señor?

¿Cómo se llama este señor?

¿Cómo se llama este señor?

-- Es el profesor

Es el profesor

Emilio González.

Emilio González.

Conoces a su sobrina

Conoces a su sobrina

que está

que está

en la misma clase.

en la misma clase.

Es el profesor Emilio González. Conoces a su sobrina que está

en la misma clase.

Es el profesor Emilio Gonzalez. Conoces a su sobrina que está

en la misma clase.

-- ¿María del Carmen?

¿María del Carmen?

Pero ella no es mexicana.

Pero ella no es mexicana.

¿María del Carmen? Pero ella no es mexicana.

¿María del Carmen? Pero ella no es mexicana.

-- No,

No,

la familia es de Panamá,

la familia es de Panamá,

pero don Emilio

pero don Emilio

es profesor

es profesor

en la Universidad de México.

en la Universidad de México.

No, la familia es de Panamá, pero don Emilio es profesor en la Universidad de México.

No, la familia es de Panamá, pero don Emilio es profesor en la Universidad de México.

— Bueno,

Bueno,

como no estuve en la clase,

como no estuve en la clase,

voy a pedir

voy a pedir

a María del Carmen

a María del Carmen

que me explique

que me explique

el arte azteca.

el arte azteca.

Bueno, como no estuve en la clase, voy a pedir a María del Carmen

que me explique el arte azteca.

Bueno, como no estuve en la clase, voy a pedir a María del Carmen

que me explique el arte azteca.

(Each line is spoken twice, followed by a pausa for repetition.)

-- ¿Qué te pasó ayer? No viniste a la Universidad.

¿Qué te pasó ayer? No viniste a la Universidad. (repetition)

-- No pude. Tuve un mal día. Estuve enfermo.

No pude. Tuve un mal día. Estuve enfermo.

-- ¡Qué lástima! La clase fue estupenda, y además no tuvimos el examen.

¡Qué lástima! La clase fue estupenda, y además no tuvimos el examen.

-- ¿Por qué no? ¿De qué habló el profesor?

¿Por qué no? ¿De qué habló el profesor?

-- Tampoco vino él. hue un amigo suyo, un artista.

Tampoco vino él. Fue un amigo suyo, un artista.

-- ¿Fue el profesor de arte indígena que acaba de llegar de México?

¿Fue el profesor de arte indígena que acaba de llegar de México?

-- Sí, nos dio una conferencia sobre el arte azteca.

Sí, nos dio una conferencia sobre el arte azteca.

-- ¿Cómo se llama este señor?

¿Cómo se llama este señor?

-- Es el profesor Emilio González. Conoces a su sobrina que está en la misma clase.

Es el profesor Emilio González. Conoces a su sobrina que está en la misma clase.

-- ¿ María del Carmen? Pero ella no es mexicana.

¿ María del Carmen? Pero ella no es mexicana.

-- No, la familia es de Panamá, pero don Emilio es profesor en la Universidad de México.

No, la familia es de Panamá, pero don Emilio es profesor en la Universidad de México.

-- Bueno, como no estuve en la clase, voy a pedir a María del Carmen que me explique el arte azteca.

Bueno, como no estuve en la clase, voy a pedir a María del Carmen que me explique el arte azteca.

6.4.1.2. The phrasing is indicated by slant lines; words used for prompting are underlined.

-- ¿Qué te pasó ayer? / No <u>viniste</u> / a la Universidad.

-- No pude. / Tuve un mal día. / Estuve <u>enfermo</u>.

-- ¡Qué lástima! / La <u>clase</u> / fue estupenda, / y además / no tuvimos el examen.

-- ¿Por qué no? / ¿De qué <u>habló</u> el profesor?

-- Tampoco vino él. / Fue un <u>amigo</u> suyo, / un artista.

-- ¿Fue el profesor / de arte indígena / que acaba de llegar / de México?

-- Si, / nos dio una <u>conferencia</u> / sobre el arte azteca.

-- ¿Cómo se <u>llama</u> / este señor?

-- Es el profesor / Emilio Gonzalez./ <u>Conoces</u> a su sobrina / que está / en la misma clase.

-- ¿María del Carmen? / Pero ella no es <u>mexicana</u>.

-- No, / la <u>familia</u> es de Panamá, / pero don Emilio / es profesor / en la Universidad de México.

-- Bueno, / como no estuve en la clase, / voy a pedir / a María del Carmen / que me <u>explique</u> / el arte azteca.

6.4.1.3. Symbols above the words indicate the items scored in pronunciation.

[i] [u]
-- ¿Qué te pasó ayer? No viniste a la Universidad.

[á] [a] [e]
-- No pude. Tuve un mal día. Estuve enfermo.

[a] [we] [´]
-- ¡Qué lástima! La clase fue estupenda, y además no tuvimos el examen.

Ø [r]
-- ¿Por qué no? De qué habló el profesor?

[y] [a]
-- Tampoco vino él. Fue un amigo suyo, un artista.

[x] [y]
-- ¿Fue el profesor de arte indígena que acaba de llegar de México?

[yo] [o][e] [rt]
-- Sí, nos dio una conferencia sobre el arte azteca.

[ñ]
-- ¿Cómo se llama ese señor?

[o][s]
-- Es el profesor Emilio González. Conoces a su sobrina que está en la misma clase.

[r] [k] [y] [x]
-- ¿María del Carmen? Pero ella no es mexicana.

[i] [a] [r] [o] [e]
-- No, la familia es de Panamá, pero don Emilio es profesor en la [d]
Universidad de México.

[we]
-- Bueno, como no estuve en la clase, voy a pedir a María del Carmen [´]
que me explique el arte azteca.

6.4.2. Dialog 2.

6.4.2.1. Spanish text.

-- Pilar, por Dios, apura. La comedia empieza a las ocho.

-- ¿Qué hora es? El teatro está muy cerca.

-- Ya son las siete y media, y además no tenemos asientos numerados.

-- Hay tiempo. No te preocupes.

-- Sí, pero es el estreno de una obra de Miguel Mihura.

-- ¡Claro! Habrá mucha gente, pero nadie llega a tiempo.

-- Pero yo quiero un asiento en la segunda fila.

-- ¡Cálmate! Estoy contigo en un momento.

-- ¿Cómo es que las mujeres tardan tanto en vestirse?

-- Es que los hombres no les ayudan nada en la casa.

-- Pilar, me imagino que ya habrá una cola de mil personas para la comedia.

-- Vamos, pues ya estoy lista.

6.4.2.2. English equivalent.

-- For heaven's sake, Pilar, hurry up. The play starts at 8:00 p.m.

-- What time is it? The theatre is very close.

-- It's already seven thirty, and besides we don't have numbered seats.

-- There is time. Don't worry.

-- Yes, but it's opening night of a play by Miguel Mihura.

-- Sure! There will be a lot of people, but nobody comes on time.

-- But I want a seat in the second row.

-- Calm down! I'll be with you in a minute.

-- Why is it that women take so long to get dressed?

-- Because men don't give them any help in the house.

-- Pilar, by now there must be a thousand people in line for the play.

-- Let's go, I'm ready.

6.4.2.3. Phrasing; prompting words.

-- Pilar, / por Dios, / apura. / La comedia / empieza a las ocho.

-- ¿Qué hora es? / El teatro / está muy cerca.

-- Ya son / las siete y media, / y además / no tenemos asientos numerados.

-- Hay tiempo. / No te preocupes.

-- Sí, / pero es el estreno / de una obra / de Miguel Mihura.

-- ¡Claro! / Habrá mucha gente, / pero nadie llega a tiempo.

-- Pero yo quiero / un asiento / en la segunda fila.

-- ¡Cálmate! / Estoy contigo / en un momento.

-- ¿Cómo es / que las mujeres / tardan tanto/ en vestirse?

-- Es que los hombres / no les ayudan nada / en la casa.

-- Pilar, / me imagino / que ya habrá una cola / de mil personas / para la comedia.

-- Vamos, / pues ya estoy lista.

6.4.2.4. Items scored for pronunciation.

[i] [yo]
-- Pilar, por Dios, apura. La comedia empieza a las ocho.

Ø [´] [a]
-- ¿Qué hora es? El teatro está muy cerca.

[ye] [ya] [e][d]
-- Ya son las siete y media, y además no tenemos asientos numerados.

[ay]t] [o][p]
-- Hay tiempo. No te preocupes.

[r] [ge] Ø
-- Sí, pero es el estreno de una obra de Miguel Mihura.

[´] [ye][y]
-- ¡Claro! Habrá mucha gente, pero nadie llega a tiempo

[u] [l]
-- Pero yo quiero un asiento en la segunda fila.

[´] [o]
-- ¡Cálmate! Estoy contigo en un momento.

[x] [mb]
-- ¿Cómo es que las mujeres tardan tanto en vestirse?

[o] [a] [e]
-- Es que los hombres no les ayudan nada en la casa.

[r] [´] [k] [i]
-- Pilar, me imagino que ya habrá una cola de mil personas para

la comedia.

[we]
-- Vamos, pues ya estoy lista.

6.4.3. Dialog 3.

6.4.3.1. Spanish text.

-- Mamá, ¿a qué hora va a estar la cena lista?

-- A las seis y media más o menos; hoy tenemos invitados.

-- ¿Ah, sí? ¿Quiénes? ¿Los conozco?

-- Algunos amigos de la oficina de tu papá y tu tío Guillermo.

-- ¿Qué hay de cena? ¿Algo especial?

-- Pollo al horno y puré de papas con mantequilla.

-- ¿No hay verduras para acompañar?

-- Sí, ensalada de lechuga con aguacate.

-- ¡Qué rico! ¿Y mangos con crema batida de postre.

-- No, de postre tenemos torta de chocolate con helado.

-- ¡Helado! Eso me gusta mucho.

-- Hija, ¿puedes ayudarme a preparar el café?

-- Sí, mamá, ¿quieres que te ayude en algo más?

-- Sí, por favor, fíjate que hayan flores en la mesa.

6.4.3.2. English equivalent.

-- Mom, at what time will dinner be ready?

-- At about six thirty, we're having guests today.

-- Oh! Who? Do I know them?

-- Some friends from Dad's office, and uncle Bill.

-- What's for dinner, something special?

-- Roast chicken and mashed potatoes with butter.

-- Are there vegetables with it?

-- Yes, lettuce and avocado salad.

-- Delicious! And mangos with whipped cream for dessert.

-- No, for dessert we have chocolate cake and ice cream.

-- Ice cream! I like that very much!

-- Dear, could you help me make the coffee?

-- Yes, Mom, do you want me to help in anything else?

-- Yes, please see that there are flowers on the table.

6.4.3.3. Phrasing; prompting words.

-- Mamá/ ¿a qué hora va a estar / la cena lista?

-- A las seis y media / más o menos; / hoy tenemos invitados.

-- ¿Ah, sí? / ¿Quienes? / ¿Los conozco?

-- Algunos amigos / de la oficina de tu papá / y tu tío Guillermo.

-- ¿Qué hay de cena? / ¿Algo especial?

-- Pollo al horno / y puré de papas / con mantequilla.

-- ¿No hay verduras / para acompañar?

-- Sí, / ensalada de lechuga / con aguacate.

-- ¡Qué rico! / ¿Y mangos con crema batida / de postre?

-- No, / de postre tenemos / torta de chocolate / con helado.

-- ¡Helado! / Eso me gusta mucho.

-- Hija, / ¿puedes ayudarme / a preparar el café?

-- Sí mamá, / ¿quieres que te ayude / en algo más?

-- Sí, / por favor, / fíjate / que hayan flores / en la mesa.

6.4.3.4. Items scored for pronunciation

[´] Ø
-- Mamá, ¿a qué hora va a estar la cena lista?

[ey] [oi] [imb]
-- A las seis y media más o menos; hoi tenemos invitados.

[ye] [o]
-- ¿Ah, si? ¿Quiénes? ¿Los conozco?

[o] [p] [t] [y]
-- Algunos amigos de la oficina de tu papá y tu tío Guillermo.

Ø [a] [l]
-- ¿Qué hay de cena? ¿Algo especial?

[y] [u] [´] [i]
-- Pollo al horno y puré de papas con mantequilla.

[r] [ñ]
-- ¿No hay verduras para acompañar?

[d] [č]
-- Sí, ensalada de lechuga con aguacate.

[r̃] [ḡ] [tr]
-- ¡Qué rico! ¿Y mangos con crema batida de postre?

[rt] [č]o]
-- No, de postre tenemos torta de chocolate con helado.

[a]
-- ¡Helado! Eso me gusta mucho.

Ø [we] [a]
-- Hija, ¿puedes ayudarme a preparar el café?

[k]
-- Sí mamá, ¿quieres que te ayude en algo más?

[´] [e]
-- Sí, por favor, fíjate que hayan flores en la mesa.

6.4.4. Dialog 4.

6.4.4.1. Spanish text.

-- Hola Juan, ¿Cuándo llega tu familia de sus vacaciones de verano?

-- Mi papá y mi mamá llegan mañana por la tarde.

-- ¿Vienen tus hermanos también o se van a quedar hasta el fin del verano?

-- Ellos vienen después, tienen amigos en Acapulco.

-- ¿Tienen casa en la playa o se alojan en un hotel?

-- Tienen una hermosa casa de veraneo en la playa y también hay un bote.

-- Y tú, ¿vas a la playa con frecuencia?

-- ¡Claro! aunque algunas veces tendría que quedarme a estudiar.

-- Tú estudias mucho durante la semana y mereces descansar.

-- Cuando quieras puedes venir conmigo, juntos nos divertiremos mucho.

-- ¿De veras me invitas? Pues voy con mucho gusto.

-- ¿Quieres venir el próximo fin de semana?

-- Si, muchas gracias, ¿nos vemos el sábado entonces?

6.4.4.2. English equivalent.

-- Hi, John. When is your family coming back from their summer vacation?

-- My father and mother are arriving tomorrow afternoon.

-- Are your brothers coming too, or will they stay until the end of the summer?

-- They are coming later; they have friends in Acapulco.

-- Do they have a house at the beach or are they staying in a hotel?

-- They have a beautiful summer home at the beach, and there's a boat, too.

-- And do you go to the beach often?

-- Oh yes, even though sometimes I should stay here to study.

-- You study a lot during the week and deserve a rest.

-- When you like, you can come with me. We'll have a good time together.

-- Are you really inviting me? I'd love to go.

-- Do you want to go next week-end?

-- Yes, thank you very much. I'll see you Saturday then.

6.4.4.3 Phrasing; prompting words.

-- Hola Juan, / ¿Cuándo llega tu familia / de sus vacaciones de verano?

-- Mi papá y mi mamá / llegan mañana / por la tarde.

-- ¿Vienen tus hermanos también / o se van a quedar / hasta el fin del verano?

-- Ellos vienen después, / tienen amigos / en Acapulco.

-- ¿Tienen casa / en la playa / o so alojan / en un hotel?

-- Tienen / una hermosa casa de veraneo / en la playa / y también / hay un bote.

-- Y tú / ¿vas a la playa / con frecuencia?

-- ¡Claro! / aunque algunas veces / tendría que quedarme / a estudiar.

-- Tú estudias mucho / durante la semana / y mereces descansar.

-- Cuando quieras / puedes venir conmigo, / juntos / nos diverteremos mucho.

-- ¿De veras me invitas? / Pues voy / con mucho gusto.

-- ¿Quieres venir / el próximo fin de semana?

-- Si, / muchas gracias, / ¿nos vemos el sábado entonces?

6.4.4.4. Items scored for pronunciation.

[x] [y] [s]
-- Hola Juan, ¿Cuándo llega tu familia de sus vacaciones de verano?

[p] [´] [ñ]
-- Mi papá y mi mamá llegan mañana por la tarde.

Ø [i]
-- ¿Vienen tus hermanos también o se van a quedar hasta el fin del verano?

[y] [u]
-- Ellos vienen después, tienen amigos en Acapulco.

[a] [x] [l]
-- ¿Tienen casa en la playa o se alojan en un hotel?

[ye] [e] [y]
-- Tienen una hermosa casa de veraneo en la playa y también un bote.

[t]
-- Y tú, ¿vas a la playa con frecuencia?

[a] [aw] [l] [k] [r]
-- ¡Claro! aunque algunas veces tendría que quedarme a estudiar.

[u] [e] [r]
-- Tú estudias mucho durante la semana y mereces descansar.

[a] [x] [i]
-- Cuando quieras puedes venir conmigo; juntos nos divertiremos mucho.

[mb] [u]
-- ¿De veras me invitas? Pues voy con mucho gusto.

[o]
-- ¿Quieres venir el próximo fin de semana?

[a] [s] [´]
-- Si, muchas gracias. ¿Nos vemos el sábado entonces?

6.4.5. Dialog 5.

6.4.5.1. Spanish text.

-- Me dicen que viaja usted a los Estados Unidos.

-- Imagínese la dicha que tengo. Acabo de ganarme una beca.

-- Y viaja usted sola o con un grupo de estudiantes?

-- Sola; precisamente ése es el problema que tengo.

-- Pero, por qué? Es la primera vez que usted viajará en avión?

-- No, es mi familia que no quiere que me vaya sola.

-- Pero, si es solamente por un año, y le convendrá mucho para su carrera.

-- Claro que sí, pero usted sabe como son las costumbres latinas.

-- Creo que hay que sacrificarse un poco en un caso como éste.

-- Eso sí pero es difícil convencer a mi mamá.

-- De todas maneras, todavía tiene tiempo para decidrlo.

-- Sí, menos mal que tengo tres meses para prepararme.

-- Estoy segura que todo va a salir bien, buena suerte.

6.4.5.2. English equivalent.

-- They tell me you are going to the United States.

-- You can imagine how happy I am. I just won a scholarship.

-- And are you traveling alone or with a group of students?

-- Alone. That's precisely my problem.

-- But why? Is it the first time that you'll travel by plane?

-- No. It's that my family doesn't want me to go alone.

-- But if it's only for one year... and it is so good for your career.

-- Of course it is. But you know how Latin customs are.

-- I think you have to sacrifice a little in a case like this.

-- Yes, but it's hard to convince my mother.

-- Anyway, you still have time to decide.

-- Yes, it's good that I have three months to get ready.

-- I'm sure everything will turn out all right. Good luck.

6.4.5.3. Phrasing; prompting words.

-- Me dicen / que viaja usted / a los Estados Unidos.

-- Imagínese / la dicha que tengo. / Acabo de ganarme / una beca.

-- Y viaja usted sola / o con un grupo / de estudiantes?

-- Sola; / precisamente / ése es el problema / que tengo.

-- Pero, por qué? / Es la primera vez / que usted viajará / en avión?

-- No, / es mi familia / que no quiere / que me vaya sola.

-- Pero, / si es solamente por un año, / y le convendrá mucho /para su carrera.

-- Claro que sí, / pero usted sabe / como son / las costumbres latinas.

-- Creo / que hay que sacrificarse / un poco / en un caso como éste.

-- Eso sí, / pero es difícil / convencer a mi mamá.

-- De todas manera, / todavía tiene tiempo / para decidirlo.

-- Sí, menos mal / que tengo tres meses / para prepararme.

-- Estoy segura / que todo va a salir bien, / buena suerte.

6.4.5.4. Items scored for pronunciation.

[e] [ya] [d]
-- Me dicen que viaja usted a los Estados Unidos.

[x] [č] [a]
-- Imagínese la dicha que tengo. Acabo de ganarme una beca.

[x] [o] [ya]
-- ¿Y viaja usted sola o con un grupo de estudiantes?

[a] [e] [e]
-- Sola; precisamente ése es el problema que tengo.

[r] [k] [´] [a]
-- Pero, por qué? ¿Es la primera vez que usted viajará en avión?

[ye] [y]
-- No, es mi familia que no quiere que me vaya sola.

[ñ] [dr]
-- Pero, si es solamente por un año y le convendrá mucho para su
[r̄]
carrera.

[o]
-- Claro que sí, pero usted sabe como son las costumbres latinas.

ø [k] [mp]
-- Creo que hay que sacrificarse un poco en un caso como éste.

[i] [´]
-- Eso sí, pero es difícil convencer a mi mamá

[d̄] [ye]
-- De todas maneras, todavía tiene tiempo para decidirlo.

[tr] [r]
-- Sí, menos mal que tengo tres meses para prepararme.

[u] [we]
-- Estoy segura que todo va a salir bien, buena suerte.

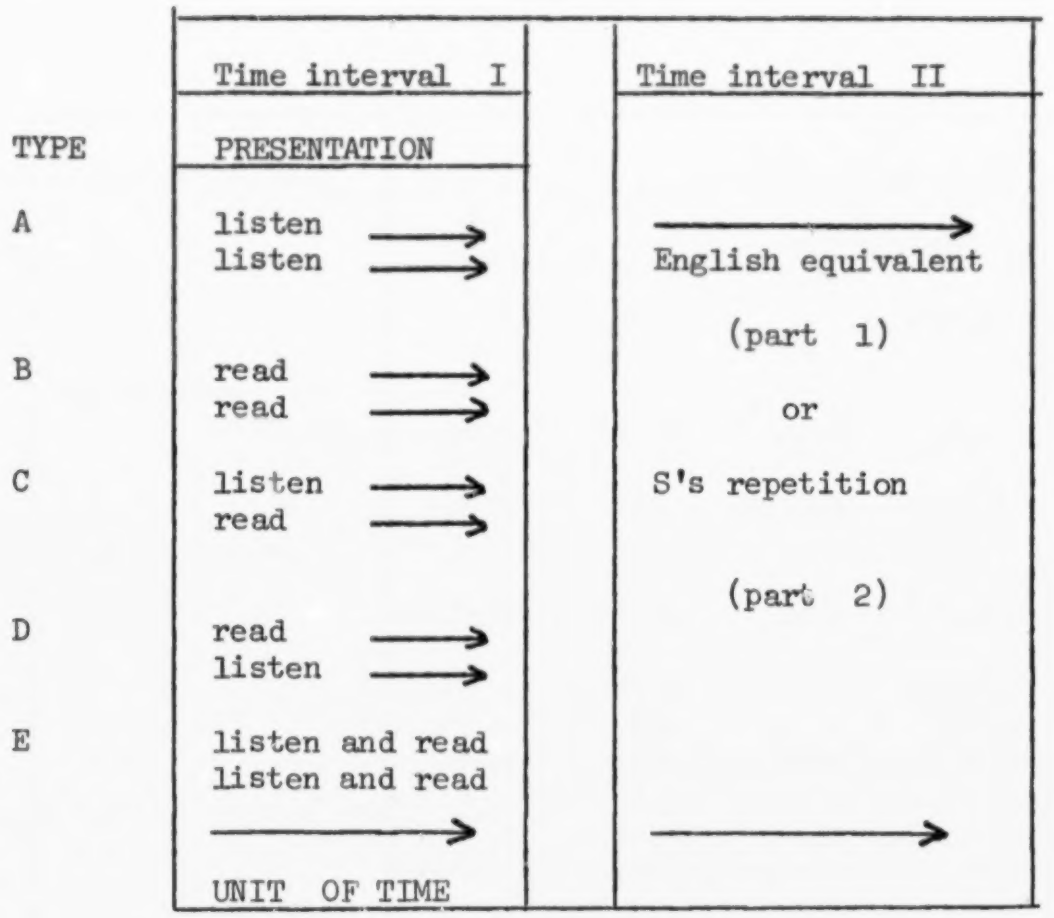
6.5. Time-sequence relationships.

The time for each visual presentation was determined by the time taken to present the same sequence at normal speaking speed. In so far as it was possible without the precision of the AV. P.A.D. synchronizer originally planned for use in the experiment, the time of the visual stimuli corresponded to the time of the audio stimuli (see 6.2.).

6.5.1. Presentation of a dialog as a whole.

TYPE	Time interval I PRESENTATION		Time interval II
A	listen → listen →		English equivalent
B	read → read →		English equivalent
C	listen → read →		English equivalent
D	read → listen →		English equivalent
E	listen and read listen and read		English equivalent
	→		→
	UNIT OF TIME		

6.5.2. Presentation of a line of a dialog.



6.5.3. Presentation of a phrase.

	Time interval I				Time interval II
TYPE	PRESENTATION				
A	listen		listen		S's repetitions
B	read	read	read	read	
C	listen	listen	read	read	→
D	read	read	listen	listen	
E	listen and read	read	listen and read	read	
	$time_1$ =	t_2 =	t_3 =	t_4	
	UNIT OF TIME →				

7. STATISTICAL ANALYSIS OF THE LANGUAGE LEARNING EXPERIMENT

OTHMAR W. WINKLER, GEORGETOWN UNIVERSITY

7.1 Summary

The experiment was based on a two-factor repeated measurements design. In much of this research we used Subjects and Type of Presentation (ToP) as the two factors.[1] In a few instances we used Subjects and Dialog, and in one instance we used Subjects, Dialog, and Type of Presentation simultaneously in a three-factor analysis of variance.

Thirty-seven test subjects learned five Spanish dialogs by five different ToPs. These were set up in the manner of a Latin square and the test subjects assigned randomly. The dialogs, however, were used in the same rigid sequence for each test subject, a most advantageous research design to study the effects of the five different ToPs. Usually groups of five subjects were submitted together to the same sequence of dialog-ToP combinations.

Analysis of variance revealed substantial and statistically significant differences between the five alternative ToPs. The observed F-ratio between the averages of written narrow unprompted (WNU) scores was F = 15, that is three times as large as the critical F-table value for a .1% level of significance[2] for four and 144 degrees of freedom (F = 4.95). With one exception, the average percentages of successfully pronounced words differed little between the five ToPs, the F-ratios for other than WNU scores were similarly substantial and significant.[3]

7.2 Description of the Statistical Model

---

[1]Because of the need to refer in the following repeatedly to "Type of Presentation," this expression will be abbreviated to ToP, its plural "Types of Presentation" to ToPs.

[2]More commonly used are the F-table values at the 1% and 5% levels of significance. These are $F_{.01} = 3.48$ and $F_{.05} = 2.45$ respectively. (For more detail see Note 11, p. 6.)

[3]These statistical terms will be explained in the next two sections, "7.2 Description of the Statistical Model" and "7.3 The Meaning of the F-Test." The reader who feels at home in this statistical terminology, however, is advised to skip these two and go directly to section "7.4 Analysis of the Principal Test Results."

The reader not fully familiar with statistical terms and assumptions will acquire a better perspective for evaluating the results if he takes the time to understand the underlying statistical model.

Each score X_{ij}[4] is assumed to be the result of the additively compounded effects of a great number of factors. But only a few of these are going to be statistically isolated and studied.

The first of these might be called the basic learning ability of the test subjects, which is assumed to be given by the arithmetic mean μ of all the elements in the population. Though underlying all computations, it is neither shown nor studied separately. In our case it is not a human population but a population of test scores, in which more than one score may have been produced by the same test subject. It is an infinite population, that is, not a set of objects or a population in the usual sense, but a process that can generate an unlimited number of scores. A complete count is not possible. Such a population can only be approached by sampling. The testing procedure employed in this language learning experiment can be considered a simple random sample.[5] The analysis of variance formulas, in fact most other statistical formulas of inference, assume that the data were collected by simple random sampling.

The second factor is the specific effect of each Type of Presentation, or ToP, on the basic learning ability in the population, α_i[6] where i = 1,2,3,4,5. Each ToP is assumed to add or subtract a fixed amount to (from) the basic score μ, and is simply determined by subtracting the grand mean μ from the mean over all the scores produced under treatment i, μ_i.

The third factor is the specific learning ability of each subject in

---

[4]The subscript i signifies a ToP and j a subject, S.

[5]There are many ways of selecting units randomly from a finite population. Simple Random requires that every unit in a finite population have the same chance of being selected, and that every set of n units be equally likely to come up. In an infinite population these conditions are not enforceable. The distinction between Simple- and other forms of random sampling does not apply in this instance and for the sake of convenience is always considered to be simple random.

[6]Read: "alpha-sub-i" or simply "alpha-i". This alpha is not to be confused with another alpha written without a subscript and used to indicate the level of significance, that is, the probability of committing an error of the first kind (Type I error), to be discussed later.

the population β_j, in which β_1 is the fixed amount added to (or subtracted from) the basic learning ability in the population due to the effect of subject 1 who according to this model is supposed to be capable of producing an infinite number of scores. β_2 is that learning ability due to subject 2, etc. There is no limit to the possible number of different β effects in the population. It is later refered to as subject effect (S).

Dialogs for most of this research were not considered as a separate factor. They were designed to differ as little as possible, and their administration was designed in such a manner that their differential effect on the scores was compensated in the aggregate by subjects or by ToPs.

All other effects are summarily treated in the model as the error component of a score. This is neither a human error nor a separate category of random causes. It is a catch-all for the many other causes which the experimenter failed to control. It is a buffer between the artificial assumptions of the model and reality. We may also consider it a waste basket which is supposed to take care of everything for which the researcher did not plan in advance. In a sense the size of this error component makes it possible to assess the quality of the researcher's conception of his problem. The smaller it is the better was his conceptualization. This error component or error variability is the remainder after subtracting α_i and β_j from the total variability of each score, $X_{ij} - \mu$ and $\varepsilon_{ij} = X_{ij} - \mu - \alpha_i - \beta_j$. Because μ., α_i and β_j are averages the error components of all the scores appear symmetrically distributed, following a normal curve at least in the idealized infinite population. It must be made clear, however, that this is a property derived from the arithmetic mean but not because the ε_{ij} is an independent random category of its own with this specific property.

According to the statistical model that underlies most of the following computations, a score in the population can be written as

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \qquad (1)$$

If this model fails to explain the scores satisfactorily by leaving too large ε_{ij} values, an interaction component $(\alpha\beta)_{ij}$ can be added to improve the model's explanatory ability in those instances.

$$X_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij} \qquad (2)$$

Model (2) is somewhat contradictorily called a nonadditive model and assumes the presence of multiplicative interaction between α_i and β_j. In our experiment that interaction can be interpreted as the idiosyncrasies of

test subjects with respect to different dialog-ToP combinations, leading to scores that cannot be explained satisfactorily by the simple additive assumption of model (1).

The repeated measurements design used for this language learning experiment is one of the most efficient experimental settings. The error component ε_{ij} however, cannot be separated from an eventually present interaction component because each subject can be tested only once in each dialog-ToP combination and only one score is available.[7] If, therefore, systematic interaction should be present, the error component in each score would be increased by the amount of interaction. This error plus interaction variability ($\sigma_{\varepsilon}^2+\sigma_{\alpha\beta}^2$) will be used in such instances as the standard of variability instead of the pure error variance which reduces the efficiency of the F-test. We will say more about this later.

These components of the model cannot be determined with certainty. They can only be estimated by sample. The degree of approximation remains essentially unknown. The basic learning ability μ is estimated by the grand mean of all sample scores $\bar{X}_{..}$[8] The effect of each type of presentation on the scores a_i, a sample-derived point-estimate of the population parameter α_i, is the difference between the average of 37 scores for ToP $\bar{X}_{i.}$[9] and the grand mean over the 185 scores, $a_i = \bar{X}_{i.} - \bar{X}_{..}$

The subject effect b_j, a point-estimate of the subject effect β_j in the population, is computed by averaging the five scores of subject j ($j = 1,2, \ldots 37$ in our sample) one score for each dialog-ToP combination. $b_j = \bar{X}_{.j} - \bar{X}_{..}$ The error variability in the sample e_{ij}, a point-estimate of ε_{ij} in the population, which may include an interaction factor, is simply the remainder after subtracting a_i and b_j from the total variability $\bar{X}_{ij} - \bar{X}_{..}$ In symbols, $e_{ij} = (X_{ij} - \bar{X}_{..}) - (\bar{X}_{i.} - \bar{X}_{..}) - (\bar{X}_{.j} - \bar{X}_{..})$

---

[7] In order to determine the error variability separately from the interaction variability, at least two scores would be needed for each test subject in each dialog-ToP combination.

[8] The dots in place of the subscripts i and j indicate the variables over which the average was taken. Here i stands for ToP effect, j for subject effect.

[9] The dot in the place of subscript j means that the scores of all the subjects who learned one of the five dialogs with ToP i have been averaged.

Each individual score in the sample can be written in analogy to (1) as

$$X_{ij} = \overline{X}_{..} + a_i + b_j + c_{ij} \tag{3}$$

or as $X_{ij} = \overline{X}_{..}+(\overline{X}_{i.}-\overline{X}_{..})+(\overline{X}_{.j}-\overline{X}_{..})+ (X_{ij}-\overline{X}_{..})-(\overline{X}_{i.}-\overline{X}_{..})-(\overline{X}_{.j}-\overline{X}_{..})$ (4)

Subtracting $\overline{X}_{..}$ on both sides gives[10]

$$X_{ij}-\overline{X}_{..}=(\overline{X}_{i.}-\overline{X}_{..})+(\overline{X}_{.j}-\overline{X}_{..})+\left[(X_{ij}-\overline{X}_{..})-(\overline{X}_{i.}-\overline{X}_{..})-(\overline{X}_{.j}-\overline{X}_{..})\right] \tag{5}$$

Squaring both sides of this equation eliminates negative signs:

$$(X_{ij}-\overline{X}_{..})^2 =(\overline{X}_{i.}-\overline{X}_{..})^2 +(\overline{X}_{.j}-\overline{X}_{..})^2 +\left[(X_{ij}-\overline{X}_{..})-(\overline{X}_{i.}-\overline{X}_{..})-(\overline{X}_{.j}-\overline{X}_{..})\right]^2 \tag{6}$$

The statistical analysis is concerned only with the larger picture, not with individual scores. The model therefore requires that all scores in the sample be aggregated. This transforms equation (6) into equation (7). In our case, the five scores of each test subject are added up first, and subsequently these 37 subtotals, one for each subject.

$$\sum_{i=1}^{5}\sum_{j=1}^{37}(X_{ij}-\overline{X}_{..})^2 = 37\sum_{i=1}^{5}(\overline{X}_{i.}-\overline{X}_{..})^2 + 5\sum_{j=1}^{37}(\overline{X}_{.j}-\overline{X}_{..})^2 +$$

$$\left[\sum_{i=1}^{5}\sum_{j=1}^{37}(X_{ij}-\overline{X}_{..})^2 - 37\sum_{i=1}^{5}(\overline{X}_{i.}-\overline{X}_{..})^2 - 5\sum_{j=1}^{37}(X_{.j}-\overline{X}_{..})^2\right] \tag{7}$$

Expression (7) can be stated simpler as an equation relating properly subscripted sum of squares:

$$SS_{Tot} = SS_T + SS_S + SS_{ST} \tag{8}$$

The last term subscripted ST (subjects within type of presentation) is the error variability and includes the variability due to the subject-ToP interaction, if present.

Table 1 gives the sums of squares and mean squares for the 185 (5x37) written unprompted narrow scores taken under the five different ToPs. The sum of squares of the three sources of variation add up to the total sum of squares. Similarly the degrees of freedom (DFs) of the three factors add up to the total number of DFs in the last row. The term degrees of

---

[10] If the parenthesis are removed and the signs adjusted, the right-hand side simplified to $X_{ij} - \overline{X}_{..}$ which proves it to be formally correct. We will also omit writing the cross-products in formula (6) as these will disappear during the process of summation in (7).

freedom essentially means sample size less the number of arithmetic means needed to determine that variability. All other influences not controlled by the experimenter are lumped together as SS_{ST}, the error of this experiment. The variability between the five ToPs is measured by SS_T. When divided by the corresponding DF (here 5 - 1 = 4) it becomes MS_T the mean square variability of ToPs.

Summarizing this discussion, we neither know nor care how the scores really come about. Instead we accept implicitly that there are groups of forces, or factors, the effects of which we can effectively isolate in the aggregate. Eventual incongruencies with reality are relegated to the catch-all factor, experimental or random error. The relative smallness of our SS_{ST} may be an indication that the model gives an acceptable representation of the scores in this experiment.

7.3 The Meaning of the F-Test

The ratio between the mean sum of squares of ToPs (MS_T) and the mean sum of squares of subjects within ToPs (MS_{ST}) for WNU scores is called the sample or observed F-ratio. It indicates in our experiment that the average variability among ToPs is 15 times larger than the error variability.

F-ratios are considered to be statistically significant at the $\alpha\%$ level if they are larger than the corresponding F-table values at the $\alpha\%$ level. In business and social science research the 5% level prevails, in physical science research the 1% level seems to prevail. We used the even stricter .1% level to underscore the enormity of the difference in our results.

The F-table gives a value of 4.95 for a .1% level of significance, for 4 DF in the numerator and 120 DFs in the denominator variability. At this point we can acknowledge only that the amount of foreign language text learned varied substantially and significantly depending on whether the subjects listened only, read only, or learned both ways with certain modifications in the sequence. They differ as a group. More detail will follow.[11]

---

[11] Dixon and Massey Jr. Introduction to Statistical Analysis, 3d ed. 1969, McGraw Hill. Such tables as a rule do not provide F-values for the exact number of DFs needed, such as for 144 DFs. In such cases we either used simply the F-values corresponding to the next smaller tabulated DF, or made linear interpolations of the next higher and lower tabulated values. The first alternative was quite acceptable because the F-values decrease with growing DFs. The F-table value corresponding to 144 DFs (not tabulated) must be smaller than that for 120 DFs which is reassuring in our particular circumstances.

As many users seem to be uncertain about the real meaning of the F-test, let a brief explanation precede the discussion of our results.

Assume the following experiment: a very large amount of numbers is generated to resemble our infinite population of scores. These numbers are constructed in such a manner that they cluster around an average score of 29.8 (which corresponds to the grand mean in our language learning experiment) of a normal distribution with standard deviation of sigma = 10.3 (also corresponding to our error variability). Each one of the approximately one million numbers is written on a small disc of durable material like cardboard. These discs are placed in a container and mixed thoroughly. A sample of 185 number discs is selected randomly from the container with some scooping device, divided into five equal batches of 37 discs each, and the arithmetic mean of the numbers computed for each batch. The values of these means will differ even though the five batches are from the same population. The between-batch variation (generally referred to as the treatment variance) and within-batch variation (called the error variance) in this special instance are simply different estimates of the same population variability. The ratio between the two should be approximately one. It is called an F-ratio (in honor of Sir Ronald Fisher) and its value is tabulated.

The 185 discs are then returned into the container, mixed well with the remaining numbers and another sample of 185 numbered discs drawn, randomly subdivided into five equal sized batches, within- and between-batch variability determined, the F-ratio computed and annotated. After doing this thousands of times, e.g. simulating such random draws on an electronic computer, all those F-ratios are ordered into a frequency distribution, which is called the sampling distribution of F, and the corresponding percentiles determined. The 90th percentile for example is 2 (really 1.99). It is that F-ratio which marks the point below which 90% of the F-ratios of that distribution are located. In a graph of that frequency distribution (see Figure 1) it will be to the left of the value 2 on the horizontal axis (a scale of F-ratios). Ten percent of the F-ratios in that distribution are as big or bigger than F=1.99. F-ratios as large or larger than 3.48 are bound to occur one in a hundred times. In other words, 99% of such F-ratios will be smaller than 3.48, only 1% will be as big or bigger than 3.48. This does not mean that this event will happen in the 100th sample; it could happen in the second or in the 807th sample. In the long run, however, chances will prove to be of that magnitude. One in a thousand or .1% of that sampling distribution of F-ratios will be as large or larger than 4.95. The remarkable thing about this distribution is that it is the only known case in which there should be no difference between the batch average and every F-ratio should be near one. The differences found eventually are due to the chance effects in the random selection of the numbers in the samples.

How large can F-ratios in that sampling distribution get? The F-ratio becomes zero if by coincidence the five batch averages happened to be exactly equal. This is the lower limit. If the within-treatment variability becomes very small, when the numbers selected randomly are identical within each batch, the F-ratio will get very large, even infinite if all the scores in each batch are identical. An F-ratio as large as 15.13 could happen perhaps once in 20,000 random experiments, or even less often.

The statistical tables containing percentiles of the F-distribution for the case of no difference in the population are not derived by such laborious procedures but computed with the mathematical function of the F-distribution. But our intuitive example should essentially lead to the same results.

The tabulated F-distribution does not contribute further to the analysis of our experiment. It answers only the question of how large F-ratios can reasonably (say 1% of the time) be expected to become if random sampling is used and the five ToPs made no difference in the population. The F-test cannot answer other questions, such as how large the real effect of the ToPs is in our experiment.

The F-ratio found in our experiment led to a rejection of the null-hypothesis that there is no difference in the effectiveness of the five ToPs. It is hard to believe that it was one of those very rare samples that showed substantial effects of ToPs when there really was none in the population. One may be inclined to consider this single experiment as sufficient proof that there really must be some typical difference in the ToPs which will show again in other samples (experiments).

As mentioned earlier, the F-test really deals with a population of test scores, not a population of test subjects. Notice that "subjects" appears in the model, like ToP, that is, like another systematic factor or treatment. Now a population of scores cannot exist independently of score-producing test subjects. The hypothetical infinite population of scores is linked to the characteristics of the score-producing subjects. These were recruited mostly from among available college students at Georgetown University.[12] The procedures by which test subjects were to be selected were not specified explicitly at the outset. The question, then, "To which kind of subject do our test results apply?" falls outside the intentions of this experiment and cannot be answered by our test results.

---

[12]The subjects selected represented a relatively narrow sector of Georgetown University's student population. They were persons of high general intelligence and language learning ability, relatively young, mainly from the School of Languages and Linguistics, the School of Foreign Service, and the College. Not included were students of Georgetown University's Schools of Law, Business Administration, Nursing, Dentistry, and Medicine.

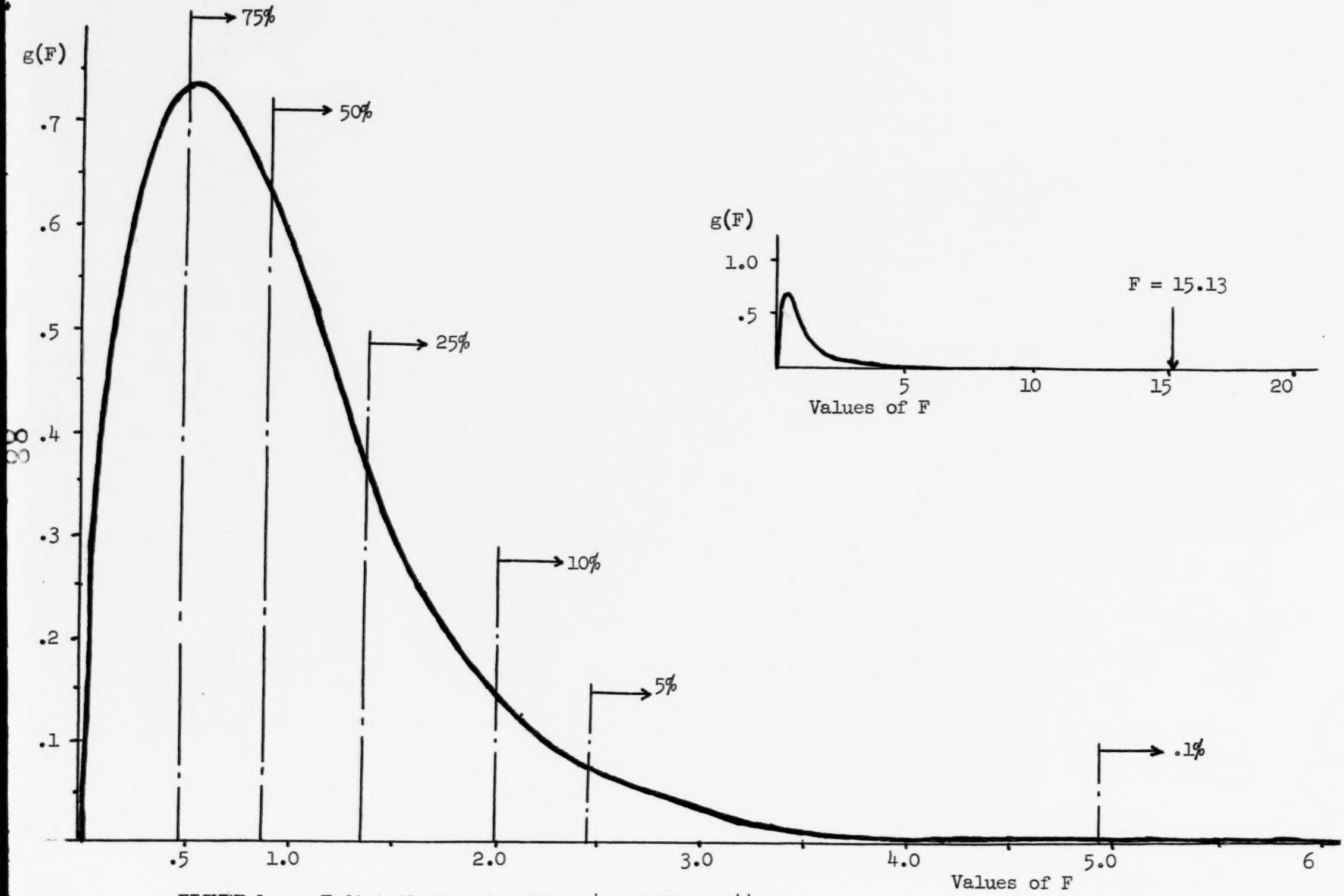


FIGURE 1 - F distribution for $DF_1 = 4$ and $DF_2 = 144$

We have no way to assess the extent to which results of our tests can be generalized to other classes of subjects. This limits the reach of our discoveries.

7.4 Analysis of the Principal Test Results

In a sense the following is a discussion of how well our experiment conforms to the statistical model of the F-test. Only if the conditions have been found to satisfy the model reasonably well can we accept the corresponding results and their implications.

A fundamental assumption of the analysis of variance model (F-test) is equality of error variabilities. The within-batch variability in each ToP is supposed to be approximately equal, that is, it must remain within the permissible range of random differences. Table 2 contains relevant information. The differences are substantial. The size of the variability obviously is associated with the size of the arithmetic means. The coefficient of correlation between standard deviations and arithmetic mean $\overline{X}$ is large, r = .753, but not significant, not even at the 5% level. A correlation coefficient in a random sample of 5 number pairs can produce freak correlations as large or larger than this in more than 5% of such random drawings even if the population of number pairs had no relationship at all (rho = 0).

Two quite different tests confirmed this. Hartley's F-test using the smallest and largest variance in the sample gives F_{max} = 729.95 / 330.20 = 2.21. The F-table value for 5 and 36 DF at the 5% level of significance is F = 2.48 which is larger than 2.21. According to Hartley's criterion, discrepancies as large as ours can obviously be expected to occur as a consequence of simple random selection and are not significant at the 5% level.

Cochran's test compares the larger variance with the sum of all the variances, in our experiment C = 729.95 / 2,434.44 = .2998. The critical value at the 5% level of significance for 5 and 36 DFs is .3066. Though close, our results are still consistent with the assumption of homogeneity of variances. Is this then a proof of homogeneity in the five variances? No. Obviously the variances differ substantially. But the tests also show that such large discrepancies must be expected more than 5% of the time if different samples of 5 batches of scores each are drawn randomly from 5 separate score populations with identical variabilities. Whatever the real situation, we must suspect that the requirement of homogeneity of variances has been violated to some extent in our experiment. It turned out to be that way, it was nobody's fault. In such a case the F-ratio tends to be larger than in experiments with homogeneous variances. Stated differently,

the probability of committing a type I error is greater.[13]

The analysis of variance model can further be violated if the covariances between each two types of presentation differ substantially. Table 3 shows the covariances between any two sets of scores by types of presentations. These covariances are between 286.2 and 544.6. Ideally, variances and covariances should be identical. If the homogeneity of variance and covariance conditions are violated, the F-ratio for MS_T/MS_{ST} (mean sum of squares between types of presentation, divided by the mean sum of squares of subjects within type of presentation) is distributed like the F-distribution, on $(a - 1)\lambda$ and $(a - 1)(n - 1)\lambda$ degrees of freedom, whereby λ approaches $1/(a - 1)$ in the extreme. Without the need for carrying out lengthy calculations for a test of homogeneity of covariances we can use the F-table on one (instead of 4) and 36 (instead of 144) DFs. The corresponding F-table values for the .1% level of significance are $F_{.001} = 12.88$, and for the more common 1% level of significance $F_{.01} = 7.41$. Both these critical values are smaller than our F = 15. We definitely can reject the null-hypothesis that there is no difference between the ToPs even if the heterogeneity of covariances were more pronounced than in our case.[14]

Another potential source of trouble is of a different nature. It was mentioned earlier that subjects were randomly assigned to pre-established test sequences. Table 4 shows that all 37 subjects were tested in each one of the five ToPs and in each one of the five dialogs. It also shows, however, that 14 of the 25 dialog-ToP combinations were used on more than 5 test subjects. To what extent did this violate the analysis of variance model? How did it affect the F-ratio?

One question that might be asked is this: can 185 scores selected randomly from a population that has the same number of scores in each one of 25 (5x5) different cells have cell frequencies as unequal as ours? We computed Chi-square assuming the observed cell frequencies as random deviations from an expected cell frequency of 7.4 scores in each dialog-ToP

---

[13]A Type I error is the mistaken conclusion that there is a substantial difference between the treatment effect, an F-ratio larger than the F-table value, when it really is only a random difference exaggerated by heterogeneity of variances.

[14]This simple method was proposed by G. E. Box, see Jerome L. Myers, Fundamentals of Experimental Design. Allyn & Bacon, 1967, 2nd print., pp. 161-162.

combination. This gave a chi-square value of 19.73[15] which lies well within the chi-square table value for the 5% level of significance of 26.296. In fact, over twenty-three percent of samples drawn from a population with equal cell frequencies in a 5x5 arrangement will lead to chi-square values equal to or larger than 19.73. We could therefore blame the uneven number of scores on random events such as equipment failure, or failure of a test subject to appear on schedule. The true reason, unfortunately, was a structural fault of our experiment. Certain dialog-ToP combinations were scheduled twice, others only once. True, the differences in the unequal number of scores are not significant. But we would commit a type II error (that is, rely on the randomness test and because our result was smaller than the critical table value, conclude "no difference" when really there was one) if we accepted this comfortable explanation.

If the dialogs were completely equal such differences would be of no concern. But they varied in length and in difficulty.

The first seemed easy to eliminate. We adjusted all scores proportionally to the length of a standard dialog of 120 words. Table 5 gives the adjustment factors (e.g. for dialog 1, that factor was 120/130 = .923077). The effect of this adjustment on the analysis of variance was minimal. The F-ratio with adjusted WNU scores, F = 15.02 differed from the F-ratio with unadjusted scores F = 15.13 by less than 1% (see tables 6 and 7). We decided, therefore, to disregard the unequal length of dialogs and used the scores without adjusting for different length of dialogs.

How important were the differences in difficulty? It soon became obvious that they were not equal, particularly dialog 2, the shortest, caused considerable learning difficulties (see table 8). Analysis of variance by dialog (see tables 7 and 9, also 11) gave an F-ratio of 10.1. Though 1/3 smaller than the F-ratios between ToPs--sum of squares for ToP (table 1, first row) is 6,392; for dialogs (table 7, first row) 4,729--it is substantial and also significant at the .1% level. Adjusted for unequal length, the F-ratios decreased from 10.1 to 8.8, obviously because part of the difference was length (number of words) of the dialogs. The error variance of the unadjusted dialogs, 117.19 (table 7 unadjusted scores, row "error", column "mean squares") is larger than the error variance of the ToPs, 105.64 (table 1, row "error", column "mean squares") because the experiment was designed to reduce error variance in ToPs, not in dialogs.

---

[15]A hypothetical cell frequency of an equal number of scores, 185/25 = 7.4 happens to be the same as an independent number of scores (37/185 x 37/185) x 185 = (.2)(.2)x185 = 7.4.

As the dialogs do differ, the fact that some dialog-ToP combinations were used more often than others requires a second, careful look at the differences which we attributed to different types of presentation. The two highest scoring ToPs B and E were tested only with 5 subjects in dialog 2, the lowest scoring ToP A was tested with 7, and ToPs C and D with 10 subjects in dialog 2. Though on the average they score higher than ToP A, there seems to be some compensation by also using the easier dialog 5 ten times each. ToPs C and D rather should be compared to the structurally related ToPs B and E (table 8). To what extent are the lower average scores due to the fact that more subjects were tested in the low scoring dialog 2? How did the easier dialogs affect the results?

To answer these disturbing questions we have rearranged the scores as shown in table 8. To adjust the score averages of each type of presentation for the unequal impact of dialogs, we first computed the average for each one of the 25 dialog-ToP cells. Then we computed new ToP average scores as if each cell had the same number of scores, by equally weighting the cell averages. These equally weighted ToP averages resemble closely those weighted proportionately (see the two bottom rows of table 8). The latter were used implicitly in all the analysis of variance computations (tables 1, 6, 13 and 14). They show a very similar picture of differences between the ToPs. True, ToP B may not have scored quite as high in the analysis of variance if 10 subjects had been tested in dialog 2 instead of only 5. But so did all other ToPs--except ToP C--and the relative differences are about the same in both kinds of averages. The net effect, then, of dialogs together with the unequal number of scores in each dialog-ToP combination need not concern us further.

As the usual analysis-of-variance procedure cannot be used because of unequal cell frequencies, the method of unweighted means (see tables 8 and 9) gives good approximations to the more laborious exact method.[16] The sum of squares and the mean squares of table 9 cannot be compared with those of tables 1, 6 or 7. The F-ratio is significant only at the 5% level because the error variance in this method is much larger than in the repeated measurements design. But this is easy to see; "error" here is the variability of each score from the cell average. Subjects with varying ability were randomly assigned to such a cell. In the repeated measurements design, "error" is the variability that remained after the subject's own average, used here as his own standard, and the average of the respective ToP had been subtracted from each score. The efficiency of this design resides exactly in the small error variability. Our experiment originally was not designed for the inefficient use made of the scores in Table 9.

---

[16]For a description of the method used, see G. W. Snedecor, Statistical Methods, 5th ed. Ames, Iowa: University Press, 1956, pp. 388-391.

There is nothing to worry about. Yet the critical reader may now wonder even more if the imperfection in our experimental lay-out--that is, using 8 instead of 5 or 10 groups of subjects--was not more related to this disappointing result than we are willing to admit.

We have tried, therefore, another way to prove that the unequal cell frequencies, though undesirable, did not really affect our results. We analyzed the variance by ToP of the first 25 test subjects. Each 5 of them went as a group exactly once through every dialog-ToP combination. The groups are identified in table 10. This arrangement neutralizes in the manner of a Latin-Square design the possible effects of differences between test subjects. If differences in the ToPs are studied, the effects of dialogs cancel out; if differences in dialogs are studied, the effects of ToPs cancel out.

Table 11 shows both, the analysis-of-variance of these 125 scores first in the 5x5 design and then in the repeated measurements design. In both instances the analysis confirmed a high level of significance between ToPs. Note that the sum of squares in both methods was 5,117. The different F-values (9.5 and 13.3) are due to the smaller error variance in the repeated measurements design (96.13 against 134.46).

Summarizing, the unequal number of scores for dialog-ToP combinations was a slight nuisance but did not affect our conclusions. This could be proved by using the first 125 scores only, which are a random subsample of the sample of 185 scores.

We referred earlier to differences in the dialogs and their effects on our study of ToPs. Part of this effect is caused neither by the length nor the structural difficulties of the dialogs, but by the rigid sequence in which the dialogs were administered: dialog 1 always first, dialog 2 next, etc. Effects such as cumulative learning belong to this category. This part of dialog variability was fully absorbed in the design and did not influence the variability between types of presentation, even though it must have influenced the F-test for dialogs and that for subjects.

Analyzing the variance in a repeated measurements design often presents the problem of nonadditivity, or interaction. Though the model equation changes from (1) to (2), nothing else really changes. Interaction cannot be noticed without special analysis of the scores. Yet its presence can make a difference. It inflates the error sum of squares, the divisor in the F-test, by an element that should not be contained in it. The F-ratios in such instances turn out to be smaller.

We applied Tukey's F-test for nonadditivity.[17] The F-ratio between the mean squares for nonadditivity with 1 DF in the numerator and mean squares of balance as the error variance with 143 DFs in the denominator was 10.7, larger than the critical F-table value at the 1% level of significance for 1 and 120 DF of 6.85 indicating the existence of a significant interaction. Table 12 shows the sum of squares which in table 1 was simply called error as error plus interaction divided into a variability due to interaction, called nonadditivity, and another component called balance which really is the pure error variability. After having satisfied ourselves that interaction of subjects with types of presentation is effectively present, we have not pursued this analysis further nor have we made attempts to adjust the F-ratios for interaction effects.[18]

A different matter that might have influenced the analysis is the loss of 3 WNU scores due to the failure of subjects 21 and 39 to appear on scheduled test sessions. That left us with only 35 complete sets. In the repeated measurement design each subject must complete all the scheduled tests. Incomplete test sequences cannot be used. There are two alternatives. One, to disregard the 7 available scores of students 21 and 39. The analysis-of-variance is carried out only with the 175 complete sets of scores of 35 test subjects. Another, more sophisticated alternative is to estimate the three missing scores (see Appendix A) and to compute the analysis of variance with the 37 sets thus completed. The 7 available test scores which had been excluded in the simpler alternative can now be salvaged, so to speak. Table 13 compares the F-ratios for ToPs, computed with 25, 26, 28, 35 and 37 sets of 5 scores each. Only the set with 37 subjects contains three estimated scores. The oral scores will be discussed later.

The steady increase in the F-ratios is due to the inclusion of scores that had been excluded in the analysis-of-variance with less subjects. This is also true of the last row (37 subjects) in which the three estimated scores may have had some additional effect. The first group of 25 subjects is not strictly comparable with the others in this table. It indicates that the inclusion of 3 estimated scores gave results consistent with the general trend visible in those sets where no estimated scores were involved. The estimation of 3 scores apparently had no influence on the F-ratios. Table 14 gives the analogous information for the F-tests on dialogs.

---

[17] See Jerome Myers, op. cit., pp. 166-169.

[18] If we used the balance mean squares of 98.95 of table 12 as the pure error variance instead of the error plus interaction variance of table 1, our F-ratio would increase from 15.13 to 16.15.

Before closing the discussion of the analysis-of-variance of WNU scores, a summary of the countervailing factors that bear on the validity of the F-test may help to clarify the real importance of our results. Considering the factors that increased and those that decreased the numerical value of the F-ratios, the following picture seems to emerge. Our experiment has violated various of the assumptions underlying the analysis-of-variance model. None of these violations were extreme. The significant interaction between subjects and ToPs increased the error component to an error plus interaction component. The real differences in the dialogs, through the somewhat lopsided design of the experiment, may further have reduced the F-ratios. On the other hand, there were balancing features that tended to increase the F-ratio, notably the heterogeneity in the variances and covariances of the five ToPs. The other features, though dismissed one by one as not sufficiently influential, may also have added effects of their own in unknown directions. Very likely the F-ratio for ToPs is not 15.13. There can be no doubt, however, that it is a large value, certainly significant at the customary levels of 5% and 1%, and probably close to F = 15.

7.5 Further Analysis of Differences between Types of Presentation

In the following section we will try to answer briefly the important question: Is the big difference between individual ToPs (see tables 8, 10) or groups of such ToPs statistically significant? Obviously ToP A was the least, ToP B the most conducive to learning. We cannot test how likely this result is to stand up in a repetition of this experiment. But we have statistical tests which tell how likely our results can be produced by coincidence (random) in a population of scores with similar characteristics but no differences between the ToPs. One is the t-test, the other Scheffe's test of multiple comparisons.

The results of t-tests between the 10 possible pairs of ToPs are shown in table 15, confirming the obvious. The differences between ToPs A and B, A and E are significant at the 1% level, between A and D at the 5% level, and between A and C at the 10% level only. No other difference was significant.

Scheffe's multiple comparisons are given in table 16. The sum of squares in column 2 were computed by

$$SS_{\psi} = \frac{\left[\Sigma w_i \left(\sum_{j=1}^{37} x_{ij}\right)\right]^2}{n \, \Sigma w_i^2}$$

where the w_i are balancing weights for the respective ToP score totals.

These weights cancel out when added.[19]

These comparisons between groups of ToP scores are considered significantly different from zero if the observed F-ratio in column 3 surpasses the adjusted F-table value $(a-1)F_{\alpha\,(a-1),a(n-1)}$ in column 6 in which F_α is the usual critical F-table value (column 5) corresponding to the alpha level of significance for (a-1) DFs, "a" being the number of compared ToPs, and (n-1) DFs, "n" being the number of scores in each ToP. An alpha level of significance of 10% is recommended as satisfactory. This covers the type I error probability for the entire group of comparisons "experimentwise," that is, the probability that at least one of the components will have a type I error.[20]

What do these results mean? There is a difference between ToP E (listening and reading simultaneously), and ToP C (listening first, then reading) in favor of ToP E significant at the 10% level, (if F-ratio in column 3 is larger than critical F-value in column 6). It must be kept in mind that the 10% level of significance is already considered as sufficient proof of the presence of statistical significance for this kind of test.

When those two, then, are contrasted as a group with ToP D (reading first, listening later) no difference remains, statistically speaking. The initial advantage of ToP E over C disappears when ToP D is brought in. These three types seem to be equivalent.

---

[19]In the comparison between ToP E and C the score total of E, 1224, is multiplied by the weight +1, the total for C, 1053, by -1.

$SS_{\psi E-C} = \frac{\left[(1)1224 + (-1)1,053\right]^2}{37(1^2 + 1^2)} = 395.15$. When comparing EC with D the weights were +1, +1 and -2. E and C are pooled here and treated like a new ToP, the average of ToP E and C.

$SS_{\psi ECD-B} = \frac{\left[(1)1224 + (1)1,053 + (1)1129 + (-3)1377\right]^2}{37(1^2 + 1^2 + 1^2 + (-3)^2)} = 1183.84$.

When comparing ECD with B the weights were +1, +1, +1 and -3, and for ECDB with A they were +1, +1, +1, +1 and -4. These comparisons are orthogonal, that is independent of one another, and without repetition. The resulting sums of squares subdivide the total sum of squares of 6392.5 (of table 1) into complementary, partial sums of squares.

[20]For a full description of this method, see Henry Scheffe, The Analysis of Variance, (New York: Wiley 1959) pp. 66-72; and Jerome L. Myers, Fundamentals of Experimental Design, (Boston: Allyn & Bacon, 1967) 2nd print., pp. 326-334.

ToPs E, C and D taken together, all containing reading and listening as essential features, contrasted strongly with ToP B (reading only) at the 2.5% level. The difference is significant, in favor of ToP B (which excludes listening from the learning process).

If E, C, D and B--which all contain reading in some form--are contrasted as a group with ToP A--the only one that relies exclusively on listening--the difference is phenomenal, in favor of the group ECDB. The advantage clearly is with the group of types that include reading in some form.

A second slightly different arrangement allows us to further explore these contrasts. The first two contrasts show nothing new, the differences being even less significant. But the contrast EDC - A adds something important. This contrast is much more significant (at the .1% level) than the contrast ECD - B considered before (significant at the 2.5% level). The group ECD, which is the same as EDC, containing listening and reading, contrasts more violently with A, listening only, than with B (just compare the sum of squares in column 2!) The conclusion suggests itself that the admixture of listening in the ToPs containing both listening and reading has a lowering effect on the scores.

The last contrast between EDCA - B confirms the significant difference in favor of reading. The two latter comparisons resulted again in F-ratios that are significant at the .1% level. This is quite unusual for multiple comparisons. The large group of EDCA is even less homogeneous as the previous contrast showed, and as a result the contrast with B is not as strong as when all reading ToPs are pitted against the only nonreading ToP. The conclusions appear clear enough. From this point of view the experiment can be called a full success.

7.6 Analysis of other Measurements

So far we discussed only WNU scores. They were the most complete information available and appeared to be reliable indicators of the amount learned by a test subject. Besides these WNU scores, 8 other scores have been made available for each subject.

Table 13 (ToPs) and table 14 (dialogs) give comparable F-ratios for various sets of WNU and ONU scores. The F-values for WNU scores increased with sample size obviously because the between ToP variability increased more than the within ToP or error variability. The conclusions drawn from a sample of 26 subjects became even more confirmed as scores from addi-

tional test subjects were (randomly) added, apart from the fact that the critical F-table values become smaller as sample size increases.

Among the other scores the ONU scores were relatively the most available. Twenty-eight subjects had complete sets of five scores each. Of the remaining subjects, subjects 1, 4, 5, 17, 19, 21, 33 and 35 lacked one ONU score each; subject 39 lacked two ONU scores. The missing scores were estimated by regressing the available ONU on the corresponding WNU scores (see Appendix A). The initial drop in the F-ratio after including those 11 subjects very likely is due to the 39 not interpolated scores that had been omitted in the analysis of variance of 28 subjects.

The F-ratio of ONU scores are also significant even at the .1% level. But they are substantially lower than those for written scores. The five types of presentation apparently are less dissimilar when measured by oral scoring.

Table 17 compares the ONU with WNU scores of all the subjects involved in the different sets of subjects in table 13, including those nine with estimated ONU scores. Some of the estimations of missing scores were not very good (e.g. for subject 1). This may have contributed to reduce the difference between the types of presentation in ONU scores, and consequently has reduced the F-ratios for ONU scores.

The F-ratio for percent pronunciation scores are not significant ($F = .73$). The number of target sounds correctly pronounced by a subject was expressed as a percentage of the total number of sounds he attempted in a given test. One successfully pronounced sound out of a total of three attempts was treated as 33%, the same as 11 successfully pronounced sounds out of a total of 33 attempts. In this way differences in basic language learning aptitude between subjects were eliminated. This type of percentage was used for the analysis of variance. In other words, the percent score averages by ToPs were computed from these simple, unweighted (or equally weighted) percent pronunciation scores. The within-treatment or error variability of the percent pronunciation scores turned out to be larger than the variability between treatments. The eleven subjects with one or more scores missing were excluded.

F-ratios smaller than 1 are infrequent. They are regarded as supporting the null hypothesis (no difference between the ToPs). If they are very small some systematic or experimenter's effect could be suspected, that is, one of the experimenters conducted his part of the experiment differently from the other experimenters, or scored the results according to somewhat divergent criteria. But the significance test for the reciprocal of this F-ratio ($F = .75$) was not significant either, indicating that no such experimenter's effect was causing our low F-ratio. All we can say here is that these differences cannot be relied on to appear

again, e.g. in the scores of another group of 26 subjects tested in the same manner, or perhaps even in the same 26 subjects tested at other times, with other dialogs. We have no assurance that the differences found among these scores by ToP represent a deeper-lying general pattern.

Accuracy of Pronunciation, Unweighted Averages

C ... 59.4% (listening first, reading afterwards)
A ... 58.8% (listening only)
D ... 57.8% (reading first, listening later)
E ... 55.5% (simultaneous listening and reading)
B ... 54.2% (reading only)

The differences between each two consecutive ToPs, at most 2.3% between D and E, are too small to allow a meaningful interpretation. They could have happened just in this particular group of subjects. Our data do not contradict the expectation that ToP A would be best, ToP B worst. To take such small differences seriously, however, we would have to get near-identical results in various repetitions of this experiment. Then we could conclude that our ranking of mean scores is more than coincidence. It would not matter then that the F-ratios were not significant in any one of them. As it stands, no conclusions should be drawn from these averages.

A different kind of average percentage results if we add the successfully pronounced words of all subjects for a given ToP and divide this sum by the corresponding total number of words attempted by those same subjects. This is like weighting each individual percentage of the kind discussed before by the total number of words attempted in each case. These averages as well as their ranking differ.

Accuracy of Pronunciation, Weighted Averages

A ... 61.9% (listening only)
E ... 57.8% (simultaneous listening and reading)
B ... 55.9% (reading only)
C ... 55.4% (listening first, reading afterwards)
D ... 55.3% (reading first, listening later)

They are a bit more spread out. ToPs B, C and D are nearly the same, A quite a bit higher. Neither does this list contradict the expectation that A would perform best, B worst. The reasons for this discrepancy between the two listings are not altogether clear. They may lie in the different manner of computing the percentages, but also in the different number of scores involved. The simple percentage-score averages were based

on 130 scores from 26 subjects. The weighted percent score averages, not used for analysis of variance, were based on 171 scores from 37 subjects. The two sets of results then represent different aspects and are not comparable to one another. Both should be taken with utter caution.

What about the other kinds of scores? After studying thoroughly the WNU scores as a prototype, it should suffice to study the degree to which the other kinds of scores followed suit. Table 18 shows F-ratios (at the margins), coefficients of correlation and coefficients of determination in the body of the table (the latter in percent) between prompted-unprompted, narrow-broad, and oral-written scores. This table summarizes what we have studied, but also what we have decided to omit. No attempts were made at estimating missing scores. The number of scores utilized are given below each ratio or coefficient in parenthesis.

The F-ratios in the 100 scores of those 20 subjects which had all the other scorings complete indeed followed closely those for the WNU scores and are, like these, highly significant. The F-ratios of the oral scores are around F = 10, those of the written scores around F = 14. As noted earlier (in table 13) the F-ratios of oral scores were consistently lower than those of the written scores.

The coefficients of correlation lie around r = .9 and are positive. "r" is an abstract measure of relationship indicating how closely the large and small WNU scores correspond to large and small scores of the other kinds. The closer this relationship, the closer to one is the value of r. The coefficient of determination is the square of the coefficient of correlation. r^2 indicates the percentage of the total variation--their direction and intensity--say in WNP scores, that coincided with that of the WNU scores. Both coefficients r and r^2 measure only the degree of association or parallelism in the scores, but do not explain it.

In the oral scores the difference between narrow-broad, prompted-unprompted are minor and may disappear if the 185 scores of all 37 subjects had been available. It should be pointed out, however, that prompting made more of a difference than broadness. ONU scores were less well associated with ONP (r = .91) and OBP scores (r = .90) than with the OBU scores (r = .96). This indicates that different kinds of scorings are more similar if both are unprompted (or prompted) than when both have the same broadness (the letters in each kind of scoring have been underlined so as to make the pairs stand out more clearly).

In the written scores the matter seemed reversed. Prompting made less of a difference--WNU were better associated with WNP scores (r=.94)

than WNU and WBU (r=.88), or WNU and WBP scores (r=.87). The scores seemed better correlated if both were of the same broadness (here: narrow) than when they were of the same prompting.

All these correlations are significant but not the differences between them. This means that the differences which we found in this experiment may not be found again in a repetition with different scores (from the same or from other subjects). But it does not mean that these differences may not be typical.

One of the first steps in this research consisted in plotting exploratory scatter diagrams (correlations charts) of the oral and written scores of this experiment with scores of various standard tests such as the Pimsleur test. As none of these scatter diagrams revealed a worthwhile relationship, analysis of covariance was discarded in favor of analysis of variance. Exact calculations later showed that analysis of covariance might have been used with modest gain with the scores of the Pimsleur 4 test (P4 in table 19). But in the meantime, analysis of variance had revealed the existence of differences in the five ToPs beyond doubt, and no sense was seen in using also analysis of covariance with the P-4 test scores.

These coefficients of simple correlation were computed after all the other promising leads had been explored. None of the coefficients were statistically significant. In other words, our sample correlations give no reason to believe that the respective coefficients of correlation in the population of scores from which we randomly had selected 185 are different from zero. The highest of these correlations coefficients, r = .41 for WNU - Pimsleur 4 scores, says only that the variability in WNU scores could be reduced on the average by about 17% when the P-4 scores for each subject had been taken into consideration.

Table 19 gives a finer breakdown of the correlation coefficients according to the ToPs. None of these simple correlation coefficients are significant, nor the differences between them. Nonetheless there seem to exist real differences between the degree in which WNU scores taken under various types of presentation associate with the various tests, particularly the ToP E and total Pimsleur score. When analyzing those coefficients it should be borne in mind that these coefficients are based only on 36 pairs of scores, as little as 1/5 the number involved in the coefficients in table 18.

APPENDIX A

We estimated three missing WNU and ten ONU scores in order to salvage the other four (in one case three) validly taken test scores of nine subjects. The ANOVA computer program only worked with complete sets of five scores for each subject. We believe that the trade-off was worth the effort and risk of such estimates. The procedures for WNU score-estimates relied on the association between scores of different subjects taking the same test sequence. The procedure for ONU scores relied on the association between the scores of the same subjects. No other personal characteristics of the subjects were taken into consideration.

Subject 21 lacked one WNU score. We plotted the scores of subjects 5, 13, 29 and 37, who had completed together with him the same test sequence of dialog-type combinations. The scores moved fairly parallel to one another. Then we applied their average change from ToP C to D (+ 43.2%) to subject 21's C score. The resulting estimate for his missing D score was 67.4. It fitted nicely into the general pattern.

Subject 39 had no B and C scores. The WNU scores of subjects 7, 15, 23 and 31 who had taken the same test sequence of dialog-type combinations with him did not have as much in common as in the previous case. But the three available WNU scores of subject 39 coincided closely with those of subjects 15 and 23. We then simply applied the percent changes of subject 23 to the available scores of subject 39, yielding score estimates of 11 and 9, which also coincided with those of subject 15.

Subject 21 and other subjects (see table 17) lacked one ONU score each. After trying alternatives, graphic linear regression of each subject's available ONU scores on his WNU scores appeared to give plausible ONU score-estimates. Subject 1 was an exception. His high WNU scores corresponded to low ONU values, and vice versa. As his missing score corresponded to a very low WNU value, the regression line-estimated ONU score gave the improbably high score of 95. He was probably not even capable of achieving such a high score, and even less in type of presentation A which generally scored lowest. This estimate happened so early in this research that its implausibility was then overlooked. It upset somewhat the general pattern and better had been omitted.

TABLE 1

Analysis of variance of 185 unadjusted written narrow unprompted (WNU) test scores, for five different types of presentation

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F-ratio
Type of Presentation	6,392.4648	4	1,598.1121	15.1285
Subjects	72,916.3750	36	2,025.4548	(19.1739)
Error	15,211.5898	144	105.6360	-
Total	94,520.3750	184	-	-

TABLE 2

Unadjusted, WNU scores, variances, standard deviations and arithmetic means

		Type of Presentation				
		A	B	C	D	E
Variance	S_j^2	330.2	730.0	448.2	495.4	430.7
Standard deviations	S_j	18.2	27.0	21.2	22.3	20.8
Arithmetic means	X_i	19.6	37.2	28.5	30.5	33.1

TABLE 3

Covariances, square roots of covariance and coefficient of correlation r between the 5 types of presentation. 185 unadjusted WNU scores

	Type of Presentation B	C	D	E
A Cov.	395.35	329.73	323.48	286.20
$\sqrt{\text{Cov.}}$	19.88	18.16	17.19	16.92
r	.81	.86	.80	.76
B Cov.		492.72	544.61	469.40
$\sqrt{\text{Cov.}}$		22.20	23.34	21.67
r		.86	.91	.84
C Cov.			390.20	336.85
$\sqrt{\text{Cov.}}$			19.75	18.35
r			.83	.77
D Cov.				350.48
$\sqrt{\text{Cov.}}$				18.72
r				.76

TABLE 4

Number of WNU test scores on which the analysis of variance was based by dialog and type of presentation combinations

	Type of presentation					Total number
	A	B	C	D	E	
Dialog 1	10	5	5	5	12	37
2	7	5	10	10	5	37
3	10	10	5	7	5	37
4	5	10	7	5	10	37
5	5	7	10	10	5	37
Total number 37	37	37	37	37	37	185

TABLE 5

Length of dialogs in number of words

Dialog	1	2	3	4	5
Number of words	130	109	133	126	133
Adjustment factor	.9231	1.1009	.9023	.9524	.9023

TABLE 6

Analysis of variance 185 WNU test scores of 37 students adjusted for differences in length of dialogs

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F-ratio
Type of presentation	6,049.85938	4	1,512.46484	15.02
Subjects	70,712.7500	36	1,964.24292	(19.51)
Error	14,498.64844	144	100.68506	-
Total	91,261.18750	184	-	-

TABLE 7

Analysis of variance of differences between averages by dialogs unadjusted and adjusted WNU scores

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F-ratio
		Unadjusted scores		
Dialog	4,729.22	4	1,182.30	10.09
Subjects	72,916.38	36	2,025.45	(17.28)
Error	16,874.85	144	117.19	1.00
Total	94,520.45	184	-	-
		Adjusted scores		
Dialog	4,062.89	4	1,015.72	8.87
Subjects	70,712.75	36	1,964.24	(17.16)
Error	16,485.65	144	114.48	1.00
Total	91,261.29	184	-	-

TABLE 8

Averages of 185 unadjusted WNU scores of all 37 students, by dialog and type of presentation (number of scores in parenthesis)

Dialog	Type of Presentation A	B	C	D	E	Equally weighted averages, by dialog
1	21.1 (10)	31.2 (5)	23.6 (5)	20.8 (5)	33.6 (12)	26.1 (37)
2	11.14 (7)	19.4 (5)	18.4 (10)	32.8 (10)	21.0 (5)	20.5 (37)
3	30.6 (10)	41.2 (10)	30.8 (5)	28.6 (7)	31.4 (5)	32.5 (37)
4	16.2 (5)	46.9 (10)	31.0 (7)	28.2 (5)	40.6 (10)	32.6 (37)
5	9.6 (5)	34.7 (7)	38.0 (10)	35.6 (10)	30.4 (5)	29.7 (37)
Equally weighted averages, by type of presentation	17.7 (37)	34.9 (37)	28.4 (37)	29.2 (37)	31.4 (37)	28.27 (185)
Proportionately weighted averages	19.6	37.2	28.5	30.5	33.1	

TABLE 9

Analysis of variance, method of unweighted means of 185 unadjusted WNU scores of Table 8

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F-ratio
Dialog	515.16	4	128.79	1.75
Type of presentation	814.40	4	203.60	2.76 ($F_{.05}$ = 2.37)
Dialog- ToP interaction	763.36	16	47.71	.65
Error (within cell variation of scores	11,809.67	160	73.81	1.00
Total	13,902.59	184	-	-

TABLE 10

Averages of the 125 unadjusted WNU scores of the first 25 test subjects, by dialog and type of presentation
(All cell averages based on 5 scores, without repetition of any dialog-type of presentation combination)

Dialog	Type of Presentation A	B	C	D	E	Average by dialog
1	16.0 1*	31.2 3*	23.6 5*	20.8 4*	30.8 2*	24.5
2	12.2 2*	19.4 1*	12.8 4*	20.4 3*	21.0 5*	17.2
3	17.6 3*	37.6 5*	30.8 1*	32.4 2*	31.4 4*	29.6
4	16.2 5*	35.4 4*	37.0 2*	28.2 1*	39.2 3*	31.2
5	9.6 4*	41.6 2*	25.8 3*	32.4 5*	30.4 1*	27.8
Average by type of presentation	14.3	32.8	26.0	26.8	30.6	26.2

*These are the numbers identifying the following groups of subjects:

Group Number	Subject Numbers
1	1, 9, 17, 25, 33
2	2, 10, 18, 26, 34
3	3, 11, 19, 27, 35
4	4, 12, 20, 28, 36
5	5, 13, 21, 29, 37

TABLE 11

Analysis of variance of the first 125 scores
Unadjusted WNU scores

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F-ratio
	As a dialog-type of presentation design with 25 cells, 5 scores in each			
Dialog	2,979.67	4	744.87	5.54
Type of presentation	5,117.07	4	1,279.27	9.51 $F_{.001}=5.31$
Dialog-type	9,614.99	16	600.94	4.47
Error (S/DT)	13,445.85	100	134.46	1.00
Total	31,157.39	124	-	-
	As a repeated measurement design			
Type of presentation	5,117.07	4	1,279.27	13.31 $F_{.001}=5.31$
Subjects	16,811.37	24	700.47	7.29
Error (type-subjects)	9,228.90	96	96.13	1.00
Total	31,157.34	124	-	-

TABLE 12

Tukey's analysis of the error term for the presence of interaction between students and types of presentation
Analysis of variance of 185 unadjusted WNU scores

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F-ratio
Error and interaction	15,211.59	144	105.64	-
Nonadditivity	1,061.43	1	1,061.43	10.73
Balance	14,150.12	143	98.95	1.00

$$SS_{\text{nonadditivity}} = \frac{\left[\sum_{j=1}^{37}\sum_{i=1}^{5} x_{ij}(\bar{x}_{.j} - \bar{x}_{..})(\bar{x}_{i.} - \bar{x}_{..})\right]^2}{\sum_{j=1}^{37}(\bar{x}_{.j} - \bar{x}_{..})^2 \sum_{i=1}^{5}(\bar{x}_{i.} - \bar{x}_{..})^2}$$

$F = 10.73 > F_{.01} = 6.85$
1 and 120 D F

TABLE 13

F-Ratio for types of presentation, adjusted and unadjusted WNU, ONU and % pronunciation scores

Number of subjects	WNU Adj'd	WNU Unadj'd	ONU Adj'd	ONU Unadj'd	% Pronunciations Unadj'd
25 (includes 1,2,3,4,5,9,10, 11, 12,13,17,18,19,20,21,25, 26,27,28,33,34,35,36,37 - these are the first 25 subjects) (4, 24 DF, $F_{.001}$=6.59, $F_{.01}$=4.22)	-	13.31 (9.51)[x]	-	4.27	-
26 (includes 3,6,7,9,11,12,13,14, 15,16,18,20,22,23,24,25,26,27,28, 29,30,31,34,36,37,38 (4, 25 DF, $F_{.001}$=6.49, $F_{.01}$=4.18)	10.22	10.39	7.40	9.32	.73
28 (includes same as in 26, plus subject numbers 2 and 10) (4, 27 DF, $F_{.001}$=6.33, $F_{.01}$=4.11)	12.75	13.27	9.72	9.74	-
35 (includes same as in 28, plus subject numbers 1,4,5,17,19,33, 35) (4, 34 DF, $F^{xx}_{.001}$=5.95, $F^{xx}_{.01}$=3.94)	13.96	14.56	6.73	6.68	-
37 (includes same as in 35, plus subject numbers 21 and 39) (4, 36 DF, $F^{xx}_{.001}$=5.86, $F^{xx}_{.01}$=3.91)	15.02	15.13	7.34	7.14	-

[x]F-ratio from a 5x5 dialog-type of presentation arrangement. Not directly comparable with other F-ratios in this table, computed from a repeated measurements design.

[xx]Linear interpolation of F-table ratios.

TABLE 14

F-ratios for dialogs, adjusted and unadjusted, WNU, ONU and % pronunciation scores

Number of subjects	WNU		ONU		% Pronunciation
	Adj'd	Unadj'd	Adj'd	Unadj'd	Unadj'd
26[x]	7.32	7.80	7.07	8.13	.98
28[x]	7.30	8.35	7.01	8.87	-
35[x]	9.19	10.47	4.00	5.73	-
37[x]	8.72	10.09	4.27	6.11	-

[x]For subjects included and critical F-table values, see first column of Table 13.

TABLE 15

t-values for pairs of differences between WNU average scores of types of presentation

	B	C	D	E
A	3.29	1.94	2.32	2.98
B	-	1.55	1.16	.74
C	-	-	.41	.95
D	-	-	-	.51

Level of significance	t-value (72DF)
.10	1.671
.05	2.000
.01	2.66

$$t = \frac{\bar{X}_A - \bar{X}_B}{S_{AB}}$$

$$S_{AB} = \sqrt{\frac{\Sigma (x_{iA} - \bar{x}_{.A}) + \Sigma (x_{iB} - \bar{x}_{.B})}{(n_A - 1) + (n_B - 1)} \times \frac{n_A + n_B}{(n_A)(n_B)}}$$

TABLE 16

Scheffe's multiple comparison analysis of variance

Source of variance contrasts between types of presentation (1)	Sum of squares (2)	Observed F-ratio (2)÷105.63603 (3)	Degrees of freedom (4)	Critical F-table value (5)	Adjusted F-table value (6)	Highest level of significance (7)
	First arrangement of orthogonal sets of contrasts					
E - C	395.15	3.74	1 and 72	$F_{.10}$, 1, 60=2.79	1F= 2.79	10%
EC - D	1.63	.02	2 and 108	$F_{.10}$, 2, 60=2.39	2F= 4.78	none
ECD - B	1,183.84	11.21	3 and 144	$F_{.025}$,3,120=3.23	3F= 9.69	2.5%
ECDB - A	4,811.85	45.55	4 and 180	$F_{.001}$,4,∞ =4.62	4F=18.48	.1%
Total	6,392.47	-	-	$F_{.001}$,4,∞ =4.62	-	-
	Second arrangement of orthogonal sets of contrasts					
E - D	121.96	1.15	1 and 72	$F_{.25}$, 1, 60=1.35	1F=1.35	none
ED - C	274.82	2.60	2 and 108	$F_{.25}$, 2, 60=1.42	2F=2.84	none
EDC - A	3,429.63	32.47	3 and 144	$F_{.001}$,3,120=5.79	3F=17.37	.1%
EDCA - B	2,566.06	24.29	4 and 180	$F_{.001}$,4,∞ =4.62	4F=18.48	.1%
Total	6,392.47	-	-	$F_{.001}$,4,∞ =4.62	-	-

TABLE 17

Comparison between unadjusted ONU and WNU scores of selected subjects

Dialog	Subject 1 Type	ONU	WNU	Subject 4 Type	ONU	WNU	Subject 5 Type	ONU	WNU	Subject 17 Type	ONU	WNU
1	A	[95]X	19	D	15	35	C	25	16	A	23	22
2	B	68	26	C	24	22	E	17	14	B	8	4
3	C	30	34	E	37	43	B	12	20	C	31	27
4	D	10	38	B	29	33	A	3	11	D	49	39
5	E	23	36	A	[7]	9	D	[31]	22	E	[38]	31

Dialog	Subject 19 Type	ONU	WNU	Subject 33 Type	ONU	WNU	Subject 35 Type	ONU	WNU	Subject 21 Type	ONU	WNU
1	B	34	42	A	15	4	B	38	37	C	42	47
2	D	14	11	B	25	22	D	16	21	E	34	34
3	A	18	9	C	32	32	A	26	35	B	56	63
4	E	[35]	45	D	[26]	26	E	74	65	A	30	27
5	C	19	14	E	24	29	C	[30]	34	D	[56]	[67]

Dialog	Subject 39 Type	ONU	WNU	Subject 2 Type	ONU	WNU	Subject 10 Type	ONU	WNU
1	E	20	15	E	79	65	E	19	17
2	D	4	5	A	26	17	A	4	8
3	A	15	9	D	58	63	D	20	25
4	B	[16]	[11]	C	55	53	C	22	19
5	C	[14]	[9]	B	75	72	B	53	46

X Boxed-in scores were estimated, for procedures see Appendix A.

TABLE 18

F-ratios, coefficients of correlation r, and coefficients of determination r^2 (in %) between unadjusted oral and written, prompted and unprompted, narrow and broad scores. The number of scores used is given in parenthesis.

		WNU	ONU	ONP	OBU	OBP	F-ratio
ONU	r	.90	1.0	.91	.96	.90	10.66
	n	(100)	(185)	(100)	(133)	(133)	(100)
	r^2	81.0%	100.0%	82.8%	92.2%	81.0%	
WNU	r	1.0	.90	-	-	-	13.34
	n	(185)	(100)				(100)
	r^2	100.0%	81.0%				
WNP	r	.94	-	-	-	-	14.11
	n	(100)					(100)
	r^2	88.4%					
WBU	r	.88	-	-	-	-	13.79
	n	(139)					(100)
	r^2	77.4%					
WBP	r	.87	-	-	-	-	-
	n	(139)					
	r^2	75.7%					
F-ratio*		-	10.66 (100)	9.86 (100)	9.85 (100)	-	$F_{.001}=5.16$** for $\alpha = .1\%$, 4 and 76DF

*Included are the scores of subjects No. 2,3,6,9,10,11,12,13,14, 18,20,22,23,25,26,28,29,30,37,38.

**Interpolated value for (5-1) DF in the numerator (between-ToP variability) and (5-1)(20-1) = 76 DF in the denominator (error or within-ToP variability)

TABLE 19

Coefficient of correlation
Pimsleur scores - unadjusted WNU scores
Number of score-pairs in parenthesis

	P_1	P_2	P_3	P_4	P_5	P_6	PT
All WNU scores	.10 (180)	.01 (180)	.26 (180)	.41 (180)	.10 (180)	.22 (180)	.31 (180)
Type A	-	-	.24 (36)	.36 (36)	-	.39 (36)	.33 (36)
Type B	-	-	.34 (36)	.47 (36)	-	.23 (36)	.35 (36)
Type C	-	-	-	.43 (36)	-	-	.21 (36)
Type D	-	-	-	.44 (36)	-	-	.26 (36)
Type E	-	-	.36 (36)	.45 (36)	-	.30 (36)	.51 (36)